6-1-2016

“Molecular Dynamics Simulation of an Engineered T4 Lysozyme Protein with Potential Nano-Biotechnological Applications”

Hadeer Elhabashy

American University in Cairo,
School of Science and Engineering,
Department of Physics

"Molecular Dynamics Simulation of an Engineered T4 Lysozyme Protein with Potential Nano-Biotechnological Applications"

A Thesis Submitted to

Physics Department

in partial fulfillment of the requirements for
the degree of Master of Science

by: Hadeer Abdel Karim ELSayed ELHabashy

Prof. Salah ELSheikh

Department of Physics, The American University in Cairo, Egypt.

Dr. Ahmed Shawky Moussa

Computer Science Department, Faculty of Computers and Information, Cairo University, Egypt.

Dr. Mohammad Sabri Yousef

Biophysics Department, Faculty of Science, Cairo University, Egypt.
Department of Physics, Southern Illinois University Edwardsville, Illinois, USA.

May/2016

Abstract

The idea of developing a Bio-molecular mechanical device has attracted the attention of many research groups due to its promising Nano and Biotechnological applications such as flow-control valves, switches and bio-sensors. T4 Lysozyme protein is an enzyme, which has been extensively studied. This protein provides a wealth of information regarding the structure function relationship of proteins at atomic resolution. An engineered variant of T4 Lysozyme has been reported to trigger a large scale trans-location of an engineered helix ($\sim$ 2nm), upon the addition of an external ligand. The design was based on the duplication of a surface helix, followed by manipulating the stability of an adjacent loop, which caused the duplicated helix to switch between two conformations. The purpose of this study is to investigate the dynamics of the engineered motion for potential Nano and Biotechnological applications. Many wild type and mutant static crystal structures of T4 Lysozyme have been shown to display a range of about 50 degrees in hinge bending motion between the N- and C-terminal domains of the protein, indicating intrinsic flexibility. The present molecular dynamics simulations detect similar motion in the engineered protein, within 100 nanosecond time scale. A preliminary mathematical model (solvent free) describing the hinge bending motion and the impact force is constructed. However, in the nanosecond time scale, the engineered triggered motion (the helical translocation) was partially detected only when bond constraints were significantly relaxed.

Dedication

This thesis is dedicated to my family.

For their endless unconditional love, support and encouragement

Acknowledgments

I would like to express my gratitude to my supervisor Prof. Salah ELSheikh who offered me this unique opportunity to study my master's degree at the American University in Cairo for his advice and encouragement throughout the course of my study.

I would like to express my sincere gratitude to my supervisor Dr. Mohammad Sabri Yousef for giving me the chance to work on one of his bright scientific ideas, and for his consistent guidance and great effort he put into training me in the field of scientific research.

I acknowledge my gratitude to Dr. Ahmed Shawky Moussa for his assistance with the computational details of simulations and for helping me getting computational resources to run my simulations via LinkSCEEM-2 project.

I owe a great debt of gratitude to Prof., Sci. Dr., Phys.-Math. Kholmirzo T. Kholmurodov, Head of the Computer Molecular Modeling sector Laboratory of Radiation Biology Joint Institute for Nuclear Research Dubna,Russia for his sincere effort in teaching me Molecular Dynamics Simulation using AMBER package and his keep in providing professional advises whenever needed. Without him, I would have had no chance to perform this kind research. He gave me the head start of what I call now " My specialty".

This work was supported by the LinkSCEEM-2 project, funded by the European Commission under the 7th Framework Programme through Capacities Research Infrastructure, INFRA-2010-1.2.3 Virtual Research Communities, Combination of Collaborative Project and Coordination and Support Actions (CP-CSA) under grant agreement no RI-261600. Preparatory Projects IDs: lspre150, lspre176 and lspre228. Production Project ID: pro14b102s1.

I want to acknowledge the assistance of Ms.Thekla Loizou from The Cyprus Institute, Cyprus in achieving the technical requirements is gratefully acknowledged.

I want to acknowledge the assistance of Mr.Waleed Lotfy and Mr.Khaled Salim from Bibliotheca Alexandrina, Egypt in achieving the technical requirements is gratefully acknowledged.

At the end, I would like to thank all those people who made this thesis possible and an unforgettable experience for me.

Contents

Chapter 1

Introduction and Literature Review

1.1 Proteins

Proteins are the working molecules of the cell. They are the most versatile biomacromolecules in living systems. They catalyze a wide range of chemical reactions, provide structural rigidity to the cell, control flow of material through membranes, regulate the concentration of metabolites, act as sensors and switches and cause motion and control gene function. These diverse tasks are performed by protein molecules constructed from monomeric building blocks called amino acids (A.A.).[1,2]

There are 20 different amino acids that make up all proteins on earth. Amino acids are molecules containing an α-Carbon (C_α) bonded to a hydrogen atom, an amine group (-NH_2), a carboxylic acid group (-COOH) and a side-chain (usually denoted as R) that varies among different amino acids as shown in figure (1.1a). The side chain is the characteristic property that distinguishes one amino acid from another and confers its chemical properties. The peptide bonds that connect two amino acids in a polymer is formed between the amine group (-NH_2) of one amino acid to the carboxylic group (-COOH) of the next in a reaction that liberates a water molecule as shown in figure (1.1b).

Three dimensional structures of proteins have evolved to carry out their functions efficiently and under precise control. These conformations are driven by a number of non-covalent interactions such as hydrogen bonding, ionic interactions, Van der Waals forces, and hydrophobic packing. The

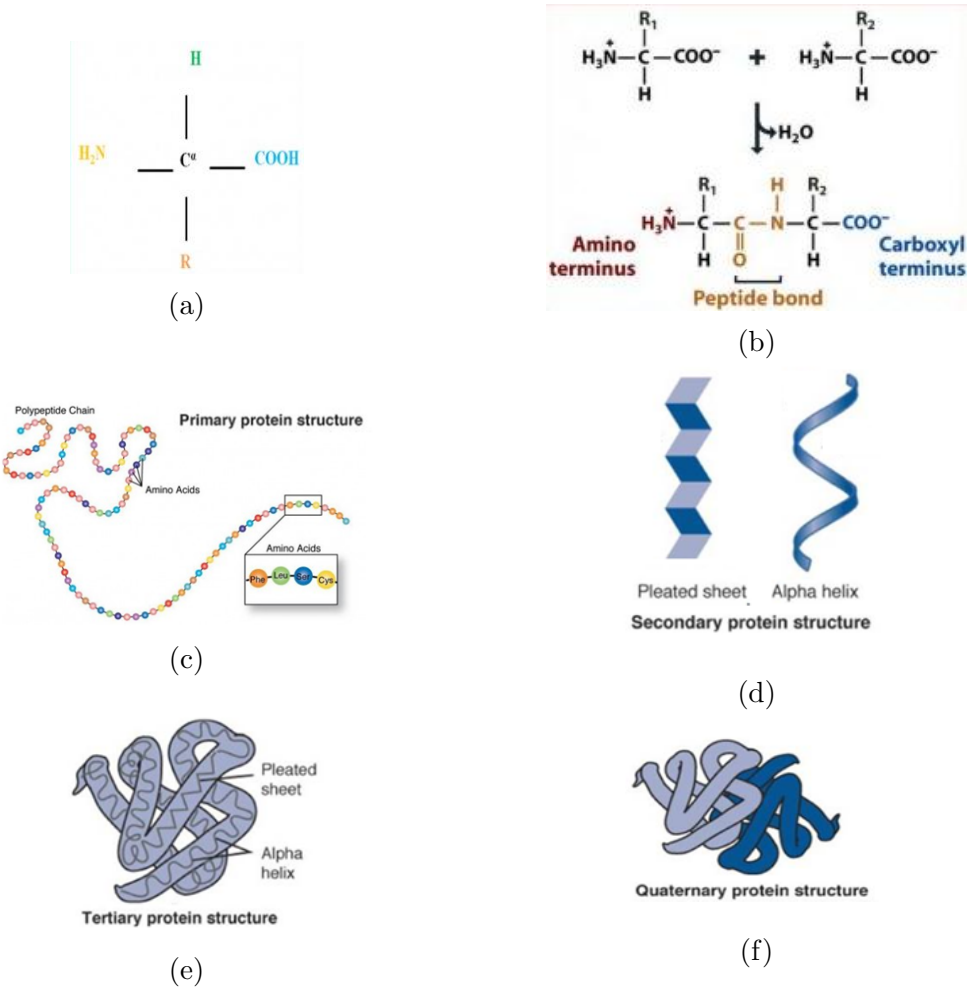


Figure 1.1: Proteins levels of organization.

(a) Amino acid general structure, (b) Peptide bond formation, (c) Primary protein structure, (d) Secondary protein structure, (e) tertiary protein structure and (f) Quaternary protein structure. (http://www.biologyexams4u.com/2011/10/protein-structure.html)

structure adapted by polypeptides can be divided into four levels of organization. Primary structure refers to the linear arrangement of amino acid residues along a poly-peptide chain and the location of the covalent bonds between chains, figure (1.1c). Secondary structure pertains to the folding of chains segments into regular structures, such as α helix and pleated sheet, figure (1.1d). Tertiary structure includes the folding of regions between α-helix and pleated β-sheet, as well as the combination of these secondary features into compact shapes or domains, figure (1.1e). Quaternary structure refers to the organization of several poly-peptide chains into a single protein molecule, figure (1.1f). The spatial organization of proteins is key to

understanding their functions.

1.2 T4 Lysozyme Protein

T4 phage Lysozyme is an enzyme produced in cells of Escherichia coli after infection with bacteriophage T4. T4 Lysozyme is a protein of 164 amino acids. Its role in the lytic cycle of the virus is to start the extra-cellular phase by liberating virions from the metabolically inert remains of the bacterial host.[3]

T4 Lysozyme has been extensively studied in literature. Early genetic and protein sequencing studies of T4 Lysozyme mutants by Streisinger and coworkers played a significant role in our present understanding of the genetic code.[4,5] Determination of the crystal structure of the protein by Matthews and coworkers set the stage for combined genetic and structural studies.[6,7] The wild-type protein, and many mutant variants, readily form crystals that diffract to high resolution. More than 200 T4 Lysozyme structures crystallized in more than 25 different crystal forms are present in the Protein Data Bank.[8] Assuming that each crystal structure represents a possible conformation in solution, this provides a unique experimental view on the conformational flexibility of the protein at atomic resolution.[9] T4 Lysozyme becomes a model enzyme in protein crystallography.[10] It has also served as a useful model system in studies of protein folding/unfolding [11] and in the development of techniques for protein nuclear magnetic resonance spectroscopy.[12]

T4 Lysozyme molecule is a promising candidate for the construction of a bio-molecular mechanical device due to its reported intrinsic conformational flexibility. The bacteriophage T4 lysozyme has long served as a model protein to study the impact of various mutations and substitutions due to its inherent tolerance to such changes.[13] Zhang et al. compared 25 non-isomorphous crystal forms of different mutants of T4 Lysozyme and found that the observed conformational flexibility, different hinge bending angles (rotational bending motions from the initial horizontal position of the duplicated helix), in different mutants of the lysozyme is an inherent property of the protein and was not a mutational effect.[14] Wray et al. examined how the effect of a surface residue mutation (L99A) is transmitted to a distantly located buried residue mutation (E108V) in T4 Lysozyme, as single residue mutations are expected to produce only localized changes in the structure.[15] Previous studies on T4 Lysozyme also suggested that it can tolerate polyalanine substitutions within the α-helices and β-sheets present in the protein. [16,17,18]

Apart from the mutational studies, insertion or relocation of different duplicate α-helical/β-sheet sequences in the wild type T4L, Matthews and coworkers have illustrated how insertion at one site can result in substantial changes at a remote site.[19–22] Sagermann et al. obtained the crystal structure of L20 mutant by inserting a duplicate α-helical sequence (residues 40i to 50i) in tandem at the N-terminal of the parent helix (40 to 50), in the wild-type protein.[19] Insertion of the duplicated helical sequence extended the parent helix by two helical turns at its N-terminus (formed by residues 43i to 50i). Whereas, the remaining inserted residues of duplicate sequence i.e., Asn40i, Ala41i and Ala42i were found to extend the N-terminal loop of the helix. The loop structure at the C-terminus of the parent helix was not perturbed in the crystal structure, owing to the interactions of Arg52 with the residues present in the C-terminal loop that preserved the parent helix at its C-terminus. [23]

Kaur and Sasidhar revealed that in the wild type T4 Lysozyme the native structure of the α helices and β sheets remained almost intact and preserved throughout 400ns of molecular dynamic simulation and the parent helical region remained helical and showed no preferences towards formation of a β-structure. While the parent helix in the mutant L20 uncoiled to form bends and turns after a series of folding /unfolding events in the final stages of the simulation. [13] They, also, examined the conformational property of an α-helical peptide sequence and its tandem duplicate in pure water and a polar environment modeled by Tetrafluoroethylene (TFE), to further understand the conformational preference of these peptides as a part of their parent protein.[24] Further, the mechanism of α to β transition as observed in duplicate peptide in pure water may have implications for folding of a rarely observed psi-loop topology. The study of the mechanism and the characterization of the observed α to β transition in the L20 mutant may provide a model system for studying α to β transition in full length protein.

An engineered model of T4 Lysozyme has been designed and reported for a new mechanical behavior, unrelated to the catalytic activity of the enzyme. A crystal structure of T4 Lysozyme is shown to prompt a large-scale translocation of the duplicated α-helix up to 20 $\mathring{A}$ upon insertion mutation by α-helix sequence duplication of 39-50 amino acids followed by a point mutation of Arginine 63 to Alanine 63 (mutant L20/R63A). [25,26] The duplicated-helix returned to its original conformation upon addition of guanidinium ligand as a replacement for the guanidinium head group of the Arginine 63 side chain. This ligand seems to control the mechanical trans-location switching between

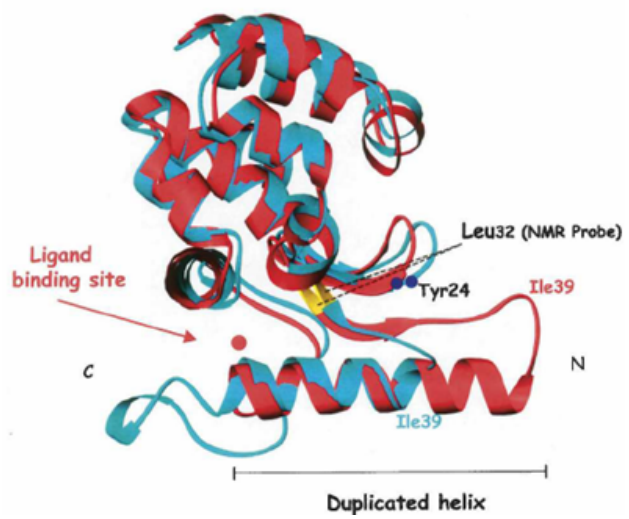


Figure 1.2: Triggered helical motion of T4 Lysozyme.

Use of a duplicated helix plus a point mutation to engineer large-scale conformational changes. The point mutation is highlighted (Arg63 to Ala in the L20/R63A mutant, corresponding to Arg52 to Ala in wild-type). Shown above is a superposition of the alter wild-type structures that can result from the use of this sequence repeat. The structures are arranged so that the duplicated helix is at the bottom of the figure with its C-terminus at the left. When guanidinium ligand binds at the ligand binding site it stabilizes the C-terminus and the helix extends to the right (ligand ed L20/R63A structure shown in red). The structure of L20, which has an Arginine at site 63, is essentially identical. In the absence of guanidinium ligand the C-terminal region of the helix is destabilized and the helix extends to the left (non-liganded L20/R63A structure shown in blue). The position occupied by Ile39 in the respective structures is shown.[26]

two alter wild-type 'on' and 'off' conformational states spanning over 20$\mathring{A}$, figure (1.2). In the "on" state, the N-terminal loop of the duplicated-helix is weakly structured, whereas the C-terminal loop has a well defined conformation that is stabilized with the Arginine 63 head group or the presence of the guanidinium ligand. In the 'off' state, the C-terminal loop became destabilized and switches the protein by the trans-location of the duplicated helix towards the N-terminal due to the absence of the Arginine 63 head group in Ala 63 or the absence of the guanidinium ligand. The introduction of the guanidinium ligand can restore the protein from the off state to on state by re stabilizing the C-terminal loop. T4 Lysozyme is a good system to study in silico, not only for the large number of available X-ray crystal structures but also because of its relatively small size which makes it suitable for computational studies.

1.3 Point of Research

The purpose of this study is to investigate, in silico, the dynamical behavior and energetics of the engineered T4 Lysozyme as a step towards the development of a bio-molecular switch. Molecular dynamics simulations coupled with experimental investigation have been routinely used to study proteins structure/function relationship and analyzing underlying interaction between molecules. In addition this study aims to build an initial mathematical model to describe motions observed by molecular dynamics simulations as a step forward to understand the physical basis and consequences of large scale intrinsic and/or triggered motions.

Chapter 2

Computational Details

2.1 Molecular Dynamics Simulation

Molecular Dynamics Simulations (MDS) can provide details concerning motions of individual particles (e.g. atoms) as a function of time. Thus, they can be used to address specific questions about the dynamic properties of a model system, often faster than experiments on the actual system. Of course, experiments play an essential role in validating the simulation methodology. Comparisons between simulation and experimental data serve to test the accuracy of the calculated results and to provide criteria for improving the methodology.

A working definition of Molecular dynamics simulation is a computer simulation of the physical movements of particles, atoms and molecules which are allowed to interact for a period of time in the context of N-particles system. A typical MD method is based on Newton's second law, or the equation of motion F=ma, where F is the force exerted on the particle, m is its mass and a is its acceleration. Forces on and between the particles and potential energy are defined by molecular mechanics force fields. Knowing the force on individual atoms, allows the determination of their accelerations. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. There are five key contents of molecular dynamics simulation, initial condition, boundary condition, force calculation, integrator/ensemble, and property calculation.

Molecular dynamic simulations is a powerful tool for the exploration of

the conformational energy accessible to molecules. Molecular dynamics simulations generate information at the microscopic level, including atomic positions and velocities. Statistical mechanics is then used to convert this to macroscopic values of interest such as pressure, energy or more specific information such as the energetics of a conformational change or the binding free energy of a particular drug. The advantage MD has over similar techniques such as Monte Carlo simulation is that MD not only allows access to the thermodynamic properties but to time-dependent phenomena too.

In molecular dynamics simulations Newton's equations of motion are numerically integrated for all atoms, which requires the evaluation of the atomic forces at each time step. The force evaluation is dominated computationally by the large number of non bonded interactions, and in particular by the long-range electrostatic interactions. Even with fast computers simplifying approximations are needed to reduce the computational time to an acceptable level. Molecular dynamics simulations can be time consuming and computationally expensive, however computers are getting faster and cheaper and the use of massively parallel machines can greatly extend the time scale and the molecular complexity that can be studied. [27]

2.2 High Performance Computing

High Performance Computing most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, or business. A supercomputer is a computer with a high-level computational capacity compared to a general-purpose computer.

Molecular dynamics simulations involve extensive calculations which demand high performance computing facilities, as well as proper software packages gaining full advantages of the given computational resources. Systems of interest are constantly growing on size and complexity which necessitates reconsideration of present algorithms not only because of the exploding calculation volumes, but also due to unsatisfactory scalability with increasing the processors number. Optimization of calculations is essential for reducing the continuance of simulation.

The effective exploitation of current high performance computing (HPC) platforms in molecular simulation relies on the ability of the present genera-

tion of parallel molecular dynamics code to make effective utilization of these platforms and their components, including CPUs and memory. [28]

2.3 Software

Simulations were performed using AMBER12 package of molecular simulation programs.[29–34] AMBER12 License owned by The American University in Cairo was used. AMBER package contains many powerful utility programs which have been used in the simulation including:

1. Leap is an X-windows-based program that provides basic model building, Amber coordinate and parameter/topology input file creation. It includes a molecular editor which allows for building residues and manipulating molecules.

2. sander is the "main" program used for molecular dynamics simulations, and is also used for replica-exchange, thermodynamic integration, and potential of mean force (PMF) calculations.

3. pmemd is an extensively-modified version of the sander program. It is faster than sander and scales better on parallel machines.

4. cpptraj is a program used to analyze MD trajectories, computing a variety of properties, like Root Mean Square Deviation from a reference structure, hydrogen bonding analysis, and so on. [34]

Visualization programs and data analysis programs includes:

1. VMD is a molecular visualization program for displaying, animating, and analyzing large bio-molecular systems using 3-D graphics and built-in scripting, (http://www.ks.uiuc.edu/Research/vmd/).[35,36]

2. PyMOL is a user-sponsored molecular visualization system on an open-source foundation, (https://www.pymol.org/). [37,38]

3. DynDom is a program to determine domains, hinge axes and hinge bending residues in proteins where two conformations are available, (http://fizz.cmp.uea.ac.uk/dyndom/). [39,40,41]

2.4 Hardware

Simulations were performed on high performance computing clusters including Cy-Tera supercomputer and Bibliotheca Alexandrina supercomputer. This work was supported by the LinkSCEEM-2 project, funded by the European Commission under the 7th Framework Programme through Capacities Research Infrastructure, INFRA-2010-1.2.3 Virtual Research Communities, Combination of Collaborative Project and Coordination and Support Actions (CP-CSA) under grant agreement no RI-261600. Preparatory Projects IDs: lspre150, lspre176 and lspre228. Production Project ID: pro14b102s1.

2.5 Scalability and Speed Tests

Speed tests have been performed on Cy-Tera supercomputer and Bibliotheca Alexandrina supercomputer. Speed of 10 ns/day of molecular dynamics simulation using 48 processors by AMBER/pmepmd software was chosen to an acceptable speed for this project. Running 100 ns of molecular dynamics simulation for 6 models has cost 6 models * 10 days * 24 hours/day * 48 processors $\simeq$ 70,000 core-hours

Chapter 3

Methodology

One can treat a molecular dynamics simulation as an experiment. Molecular dynamics simulation protocol includes sample preparation, energy minimization, heating, equilibration, production molecular dynamics run and output data analysis respectively. Figure (3.1) is a common flowchart of a regular molecular dynamics simulation. The following is a detailed description of the molecular dynamics simulation protocol used in this project.

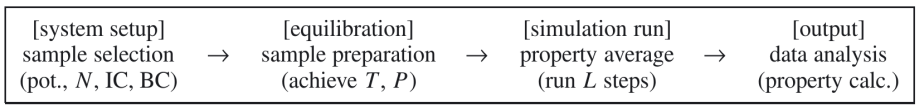


Figure 3.1: Molecular dynamics simulation common flow chart
One can often treat a MD simulation like an experiment. A common flowchart of an ordinary MD run: in which we first set up the system (here, the force field, initial conditions IC and boundary conditions BC), then fine-tune the system until it reaches the desired condition (here, temperature T and pressure P), and then perform property averages, for instance calculating the value of interest. [27]

3.1 Sample Preparation

Numerous X-ray structures of T4 Lysozyme protein are available on the protein data bank (http://www.rcsb.org/pdb).[42] Structural models with PDB IDs: 262L [19] and 2F2Q [26] were initially used. Sequences alignment between 262L and 2F2Q models is shown in figure (3.2). These models have identical structures with root mean square deviation 1.1Åwhen aligned as shown in figure (3.3). They are originally in the 'on' state, i.e. the N-terminal loop of the duplicated helix is weakly structured, whereas the C-terminal loop has a well defined conformation.

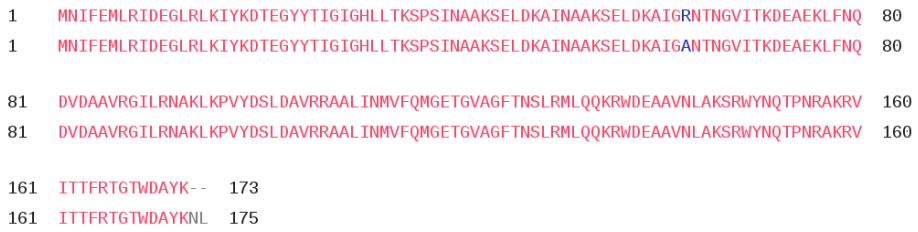


Figure 3.2: Sequence alignment between T4 Lysozyme models: 262L and 2F2Q

Sequence alignment between two models of T4 Lysozyme protein with PDB IDs 262L and 2F2Q. For model 262L, Arginine is the amino acid of site 63. For model 2F2Q, Alanine is the amino acid of site 63. Also, amino acids Asparagine 174 and leucine 175 are missing from the crystal structure of model 262L. This figure is created using the Basic Local Alignment Search Tool (Blast).

For model 262L, Arginine is the amino acid at location 63 and the C-terminal loop is stabilized by means of non-bonding interaction with guanidinium head group of Arginine 63. Model 262L has been mutated from Arginine 63 to Alanine 63 evolving a new model called mutant 262L. For model 2F2Q, Alanine is the amino acid of site 63 and C-terminal is stabilized by means of non-bonding interaction with presence of the guanidinium ligand, mimicking the Arginine side chain, to probe the structural dynamics consequences of destabilizing the C terminal loop. Guanidinium ligand was removed from the model. Four identical copies of model 2F2Q, removing the guanidinium ligand, has been created to be simulated at different condition for force field and temperature. Finally, we have 6 models including 262L, mutant 262L, one model of 2F2Q with AMBER ff99 and three models of 2F2Q with AMBER ff99SB.

AMBER ff99SB force field was used for all model except one models of 2F2Q AMBER ff99 force field was used. [43,44] A force field refers to the functional form and parameter sets used to calculate the potential energy of a system of atoms or coarse-grained particles in molecular mechanics and molecular dynamics simulations. It includes all the parameters for all of the bonds, angles, dihedrals, and atom types in the system The parameters of the energy functions can be derived from experimental work and quantum mechanical calculations.

Counter ions were added to all protein models to achieve electrical neutralization. All the crystallographic water molecules were removed from the structures. All protein models were solvated in a rectangular water box with TIP3P water model. [45] Hence, all the used water molecules in the system

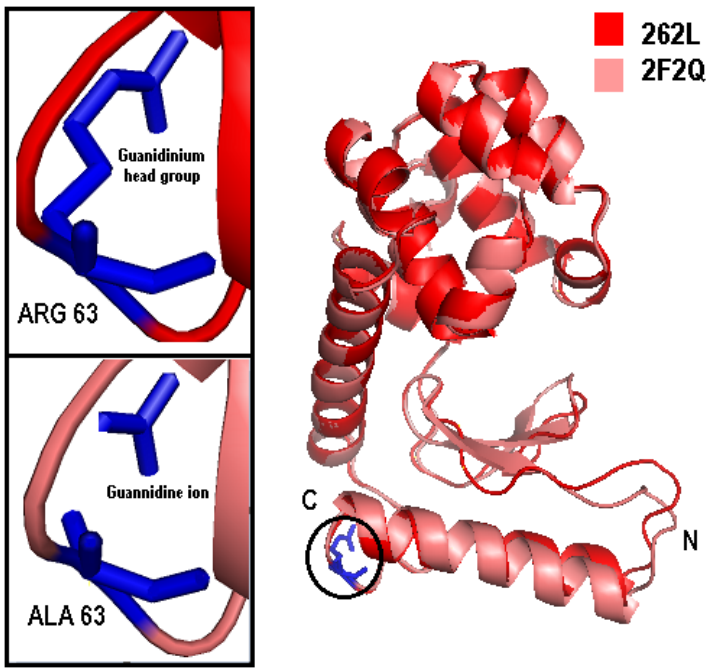


Figure 3.3: Structure alignment between T4 Lysozyme models: 262L and 2F2Q

The figure at right is showing a structure alignment between two models of T4 Lysozyme protein with PDB IDs 262L and 2F2Q. For model 262L, up left, Arginine is the amino acid of location 63 and C-terminal is stabilized by means of non-bonding interaction with guanidinium head group of Arginine 63. For model 2F2Q, down left, Alanine is the amino acid of site 63 and C-terminal is stabilized by means of non-bonding interaction with the guanidinium ligand.

will be defined by the same water model, TIP3P. Table (3.1) is showing the original conditions of the models under investigation and their molecular dynamics running conditions.

Table 3.1: The original conditions of the models under investigation and their running conditions.

Model No.	1	2	3	4	5	6
PDB ID	262L	262L	2F2Q	2F2Q	2F2Q	2F2Q
Initial state of amino acid 63	Arg	Arg	Ala	Ala	Ala	Ala
Initial state of guanidinium	Presence	Presence	Presence	Presence	Presence	Presence
Initial state of the target helix	on	on	on	on	on	on
Run state of amino acid 63	Arg	Ala	Ala	Ala	Ala	Ala
Run state of guanidinium	Presence	Absence	Absence	Absence	Absence	Absence
Run temperature	310K	310K	310K	310K	293K	277K
Force field	Amber ff99SB	Amber ff99SB	Amber ff99	Amber ff99SB	Amber ff99SB	Amber ff99SB

3.2 Energy Minimization

All models went through two rounds of minimization each of 800 steps separated by 50 ps of molecular dynamics simulation at 100K. The first round of minimization was performed only for the water molecules and the ions restraining the proteins. The weight for positional restraints was 5 $kcal/mol - \mathring{A}^2$. The second round of minimization was performed on the whole systems including the water molecules, ions and the protein without any restraints. For both rounds, 100 steps of steepest descent minimization were performed first to relieve highly unfavorable clashes followed by 700 steps of conjugate gradient minimization [46] which is much slower but more effective at reaching an energy minimum after severe clashes have been relieved. The non-bond interaction were truncated using the force shifting, where the calculated force and energies are smoothly shifted to zero at a cutoff distance of 10 $\mathring{A}$. Constant volume periodic boundary conditions were applied. SHAKE [47] was applied constraining bonds involving hydrogen. SHAKE is an algorithm based on dynamics, the minimizer is not aware of what SHAKE is doing; for this reason, minimization generally should be carried out without SHAKE. One exception is short minimization whose purpose is to remove bad contacts before dynamics begin. (See appendix .1)

3.3 System Heating

Systems heating were performed slowly at constant pressure. Models including 262L, mutant 262L, one model of 2F2Q with AMBER ff99 and one model of 2F2Q with AMBER ff99SB, were heated up from 100K to 310K. Nine rounds of molecular dynamics simulations of isothermal-isobaric ensemble (NPT) each of 50ps with temperature increment of 25K except the last round with temperature increment of 10K until a desired temperature of 310K was achieved. The other two models of 2F2Q with AMBER ff99SB were heated one up to 293K and the other up to 277K, over eight rounds and seven rounds respectively, each of 50 ps with temperature increment of 25K except the last round to achieve the target temperature.

For all rounds of heating and for the intermediate molecular dynamics simulation separating minimization rounds, the following properties were chosen and maintained. Isothermal isobaric ensemble (NPT) with langevin dynamics [48,49] was applied with collision frequency of 2 ps^{-1}. 225,000 steps were performed with step length of 0.002ps to achieve 450ps for heating up to 310K. 200,000 steps were performed with step length of 0.002ps to achieve

400ps for heating up to 293K. 150,000 steps were performed with step length of 0.002ps to achieve 350ps for heating up to 277K.

The non-bonding interactions were truncated using the force shifting, where the calculated force and energies are smoothly shifted to zero at a cutoff distance of $10\mathring{A}$. Constant pressure was maintained at 1 atm by the pressure coupling algorithms. Constant pressure periodic boundary conditions were used with isotropic pressure scaling and pressure relaxation time of 2 ps. All systems were heated restraining only the protein with 5 kcal/mol $\mathring{A}^2$ weight for the protein positional restraints. SHAKE algorithm was performed omitting all bonding interactions involving hydrogen atoms. SHAKE removes the bond stretching freedom, which is the fastest motion, and consequently allows a larger time step to be used. For water models, a special "three-point" algorithm is used. The transnational center-of-mass (COM) motion were removed at regular intervals of 500 steps. For langevin dynamics, the position of the center-of-mass of the molecule is reset to zero every 500 steps, but the velocities are not affected. Hence there is no change to either the translation or rotational components of the momenta. The only reason to even reset the coordinates is to prevent the molecule from diffusing so far away from the origin that its coordinates overflow the format used in restart or trajectory files.(See appendix .2)

3.4 Production Run

Systems were then equilibrated and simulated over for 100ns of molecular dynamics simulation. Isothermal isobaric ensembles (NPT) were used for all models closely similar to laboratory conditions of temperature and pressure. Constant temperature of 310K was maintained using Langevin dynamics with collision frequency of 2 ps-1 except for two models of 2F2Q with AMBER ff99SB temperature of 293K and 277K were maintained individually. Constant pressure periodic boundary conditions were maintained at 1bar with isotropic pressure scaling and pressure relaxation time of 2ps. 50,000,000 steps were performed with step length of 0.002ps to achieve 100ns of molecular dynamics simulations. Non-bonding interaction cutoff distance of $10\mathring{A}$ was adjusted. SHAKE algorithm was performed omitting all bond interactions involving hydrogen atoms. The translational center-of-mass (COM) motions were removed at regular interval of 500 steps. The equilibrium states of protein models were determined by the root mean square deviation of the entire protein's backbone. All simulation were performed with AMBER12 Package.(See appendix .3)

3.5 Data Analysis

Data analysis of the obtained trajectories was carried out using Amber tools provided by AMBER package.[29−34] The thermodynamics properties of each system were checked constantly along simulation time to ensure stable thermodynamic behavior. Root mean square deviation of the trajectories was calculated with respect to the initial structure using Amber tools/cpptraj program. The analysis of protein domain motion analysis was performed using the DYNDOM software (http://fizz.cmp.uea.ac.uk/dyndom/) [39,40,41]. Hydrogen bonding analysis and radial distribution functions for amino acid 63 was calculated for each model over the simulation time. Dynamical movie and snapshot of all models were created using VMD (http://www.ks.uiuc.edu/Research/vmd/) [35,36] and PYMOL software (https://www.pymol.org/) [37,38].

Chapter 4

Results and Discussion

4.1 Results

4.1.1 Initial Conformations

For 262L model and 262L mutant model at 310K with AMBER ff99SB force field, the protein crystal is originally in the 'on' state. The N-terminal loop of the duplicated helix is weakly structured, whereas the C-terminal loop has a well defined conformation. The C-terminal is stabilized by means of non-bonded interactions due to the presence of the guanidinium head group of Arginine 63. In 262L model, Arginine is originally the amino acid 63. This model is considered as a reference model. In mutant 262L model, mutation replacing Arginine 63 to Alanine 63 has been done to the structure using PYMOL software in order to probe the structural dynamics consequences of destabilizing the C-terminal loop. Simulations were performed using the updated AMBER ff99SB force field at constant temperature of 310K.

For 2F2Q model, the protein crystal is originally in the 'on' state. The N-terminal loop is weakly structured, whereas the C-terminal loop has a well defined conformation. The C-terminal is stabilized by means of non-bonded interactions due to presence of the guanidinium ligand. Removing the guanidinium ligand is done to to probe the structural dynamics consequences of destabilizing the C-terminal loop. Four molecular dynamics simulations were performed individually on four identical copies of this protein crystal under the following conditions: 1)AMBER ff99 force field and at constant temperature of 310K, 2) AMBER ff99SB force field at constant temperature of 310K, 3)AMBER ff99SB force field at constant temperature of 293K, and 4) AMBER ff99SB force field at constant temperature of 277K. Molecular dynamics simulation is going to give us a view about the dynamical behavior

of system by numerically solving Newton's equation of motion for each atom of the system.

4.1.2 Thermodynamics Properties

Thermodynamic properties of systems such as temperature, pressure, volume, density and energies were calculated each 1ps along 100ns of simulation time. The protein-water box systems were simulated as an isothermal isobaric ensemble (NPT) for 100ns of molecular dynamics simulation.

For 262L model at 310K with AMBER ff99SB force field, temperature was maintained stable around average of 310K with average Root Mean Square (RMS) fluctuations of 1.8K, figure (4.1a). pressure of 1 bar was maintained with average Root Mean Square fluctuations of 161 bar, figure (4.1b). The starting volume of the box was 286,101 $\mathring{A}^3$. However, the box volume was allowed to fluctuate through the simulation, figure (4.1c). The initial density of the system was 1 $amu/\mathring{A}^3$, figure (4.1d). However it fluctuated inversely with the volume. The total energy of the system can be decomposed into the total potential energy and the total kinetic energy, see figures (4.1f), (4.1e and 4.1g). kinetic energy of the system has shown stability with average of 17912 $kcal/mol$, since the temperature is directly proportional to the kinetic energy.

For 262L mutant model at 310K with AMBER ff99SB force field, Temperature was maintained stable around average of 310K with average Root Mean Square fluctuations of 1.8K, figure (4.2a). pressure of 1 bar was maintained with average Root Mean Square fluctuations of 156 bar, figure (4.2b). The starting volume of the box was 29,089 $\mathring{A}^3$. However, the box volume was allowed to fluctuate through the simulation, figure (4.2c). The initial density of the system was 1 $amu/\mathring{A}^3$, figure (4.2d). However it fluctuated inversely with the volume. The total energy of the system can be decomposed into the total potential energy and the total kinetic energy, see figures (4.2f), (4.2e and 4.2g). kinetic energy of the system has shown stability with average of 18,189 $kcal/mol$, since the temperature is directly proportional to the kinetic energy.

For 2F2Q model at 310K with AMBER ff99 force field, temperature was maintained stable around average of 310K with average Root Mean Square fluctuations of 1.86K, figure (4.3a). pressure of 1 bar was maintained with average Root Mean Square fluctuations of 182 bar, figure (4.3b). The start-

ing volume of the box was 258,609 $\mathring{A}^3$. However, the box volume was allowed to fluctuate through the simulation, figure (4.3c). The initial density of the system was 1 $amu/\mathring{A}^3$, figure (4.3d). However it fluctuated inversely with the volume. The total energy of the system can be decomposed into the total potential energy and the total kinetic energy, see figures (4.3f), figure (4.3e and 4.3g). kinetic energy of the system has shown stability with average of 16,260 $kcal/mol$, since the temperature is directly proportional to the kinetic energy.

For 2F2Q model at 310K with AMBER ff99 force field, temperature was maintained stable around average of 310K with average Root Mean Square fluctuations of 1.8K, figure (4.4a). pressure of 1 bar was maintained with average Root Mean Square fluctuations of 156 bar, figure (4.4b). The starting volume of the box was 290,89 $\mathring{A}^3$. However, the box volume was allowed to fluctuate through the simulation, figure (4.4c). The initial density of the system was 1 $amu/\mathring{A}^3$, figure (4.4d). However it fluctuated inversely with the volume. The total energy of the system can be decomposed into the total potential energy and the total kinetic energy,see figures (4.4f), (4.4e and 4.4g). kinetic energy of the system has shown stability with average of 18,189 $kcal/mol$, since the temperature is directly proportional to the kinetic energy.

For 2F2Q model at 293K with AMBER ff99SB force field, temperature was maintained stable around average of 293K with average Root Mean Square fluctuations of 1.7K, figure (4.5a). pressure of 1 bar was maintained with average Root Mean Square fluctuations of 160 bar, figure (4.5b). The starting volume of the box was 280,310 $\mathring{A}^3$. However, the box volume was allowed to fluctuate through the simulation, figure (4.5c). The initial density of the system was 1 $amu/\mathring{A}^3$, figure (4.5d). However it fluctuated inversely with the volume. The total energy of the system can be decomposed into the total potential energy and the total kinetic energy, see figures (4.5f), (4.5e and 4.5g).. kinetic energy of the system has shown stability with average of 16,855 $kcal/mol$, since the temperature is directly proportional to the kinetic energy.

For 2F2Q model at 277K with AMBER ff99SB force field, Temperature was maintained stable around average of 277.01K with average Root Mean Square fluctuations of 1.61K, figure (4.6a). Pressure of 1 bar was maintained with average Root Mean Square fluctuations of 156.53 bar, figure (4.6b). The starting volume of the box was 277303.1 $\mathring{A}^3$. However, the box volume was allowed to fluctuate through the simulation, see figures (4.6c), (4.6e and 4.6g). The initial density of the system was 1.0344 $amu/\mathring{A}^3$, figure (4.6d). However it fluctuated inversely with the volume. The total energy of the

system can be decomposed into the total potential energy and the total kinetic energy. kinetic energy of the system has shown stability with average of 15933.90 $kcal/mol$, since the temperature is directly proportional to the kinetic energy, figure (4.6f).

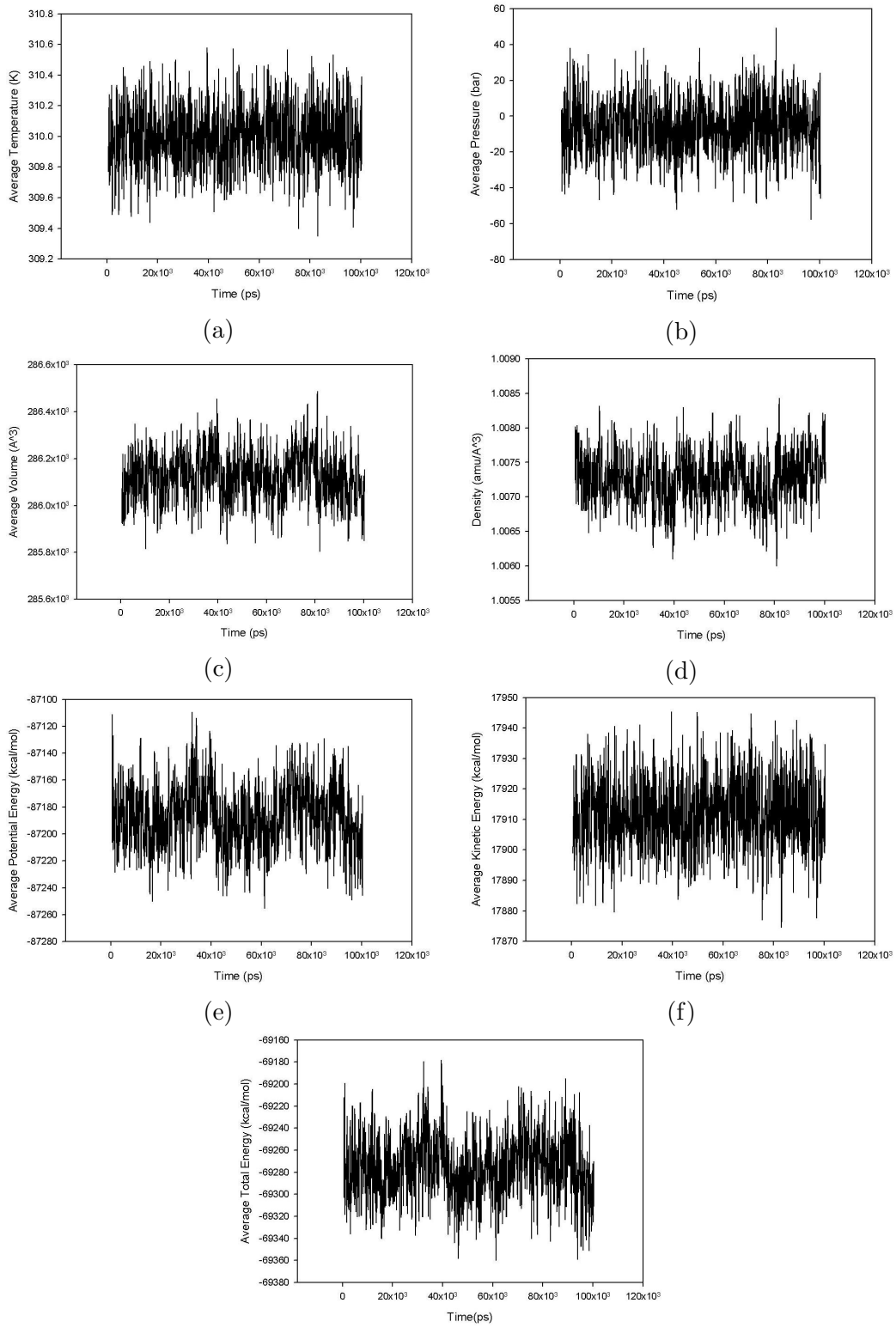


Figure 4.1: Thermodynamic properties of 262L model at 310K and ff99SB

The thermodynamic properties of 262L model of T4 lysozyme protein vs time in (ps) during 100 ns of molecular dynamics simulation including a) Average temperature (K), b) Average pressure (bar), c) Average volume ($\AA^3$), d) Average density ($amu/\AA^3$), e) Average potential energy (kcal/mol), f) Average kinetic energy (kcal/mol) and g) Average total energy (kcal/mol).

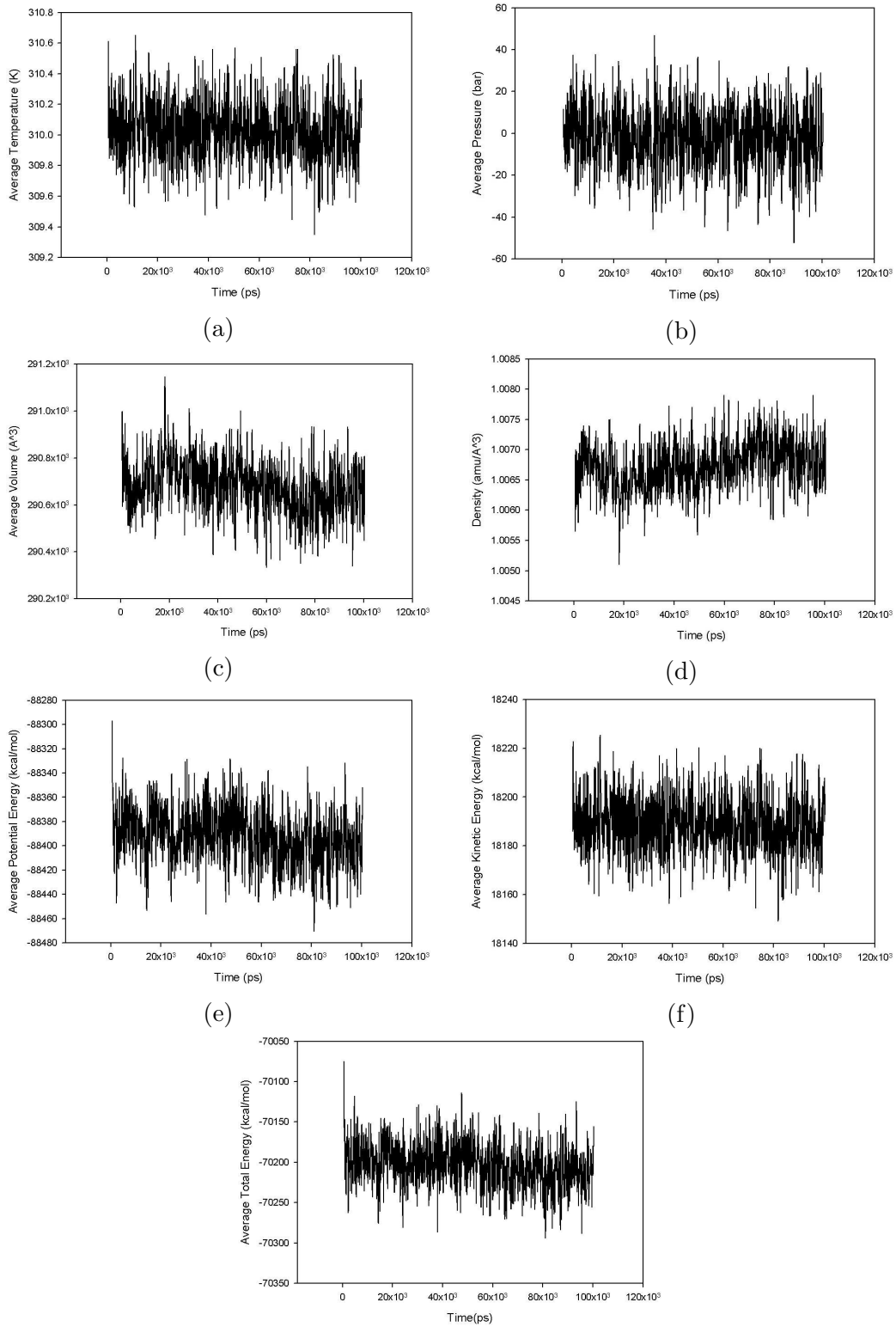


Figure 4.2: Thermodynamic properties of 262L mutant model at 310K and ff99SB

The thermodynamic properties of 262L mutant model of T4 Lysozyme protein vs time in (ps) during 100 ns of molecular dynamics simulation including a) Average temperature (K), b) Average pressure (bar), c) Average volume ($\AA^3$), d) Average density ($amu/\AA^3$), e) Average potential energy (kcal/mol), f) Average kinetic energy (kcal/mol) and g) Average total energy (kcal/mol).

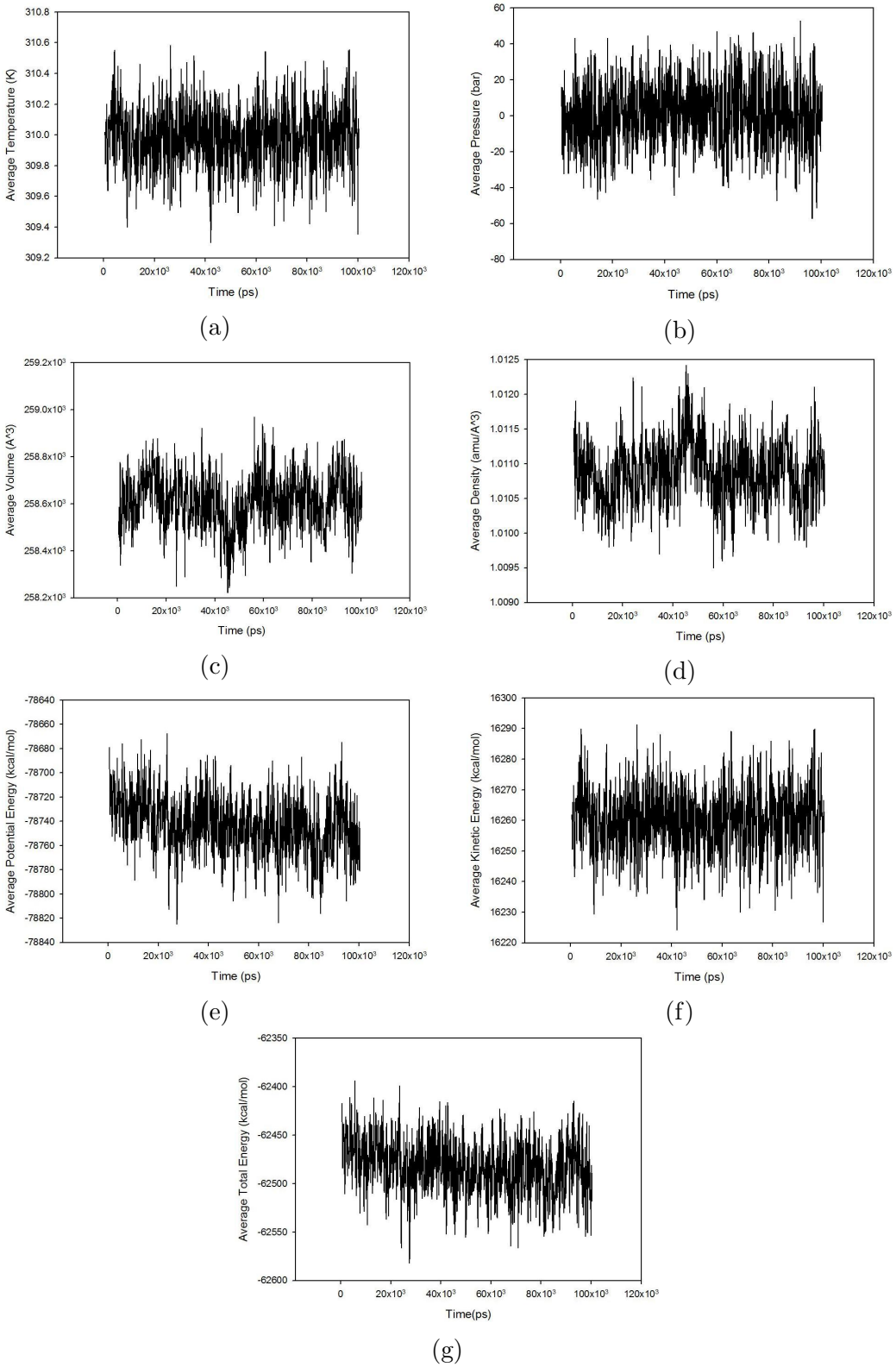


Figure 4.3: Thermodynamic properties of 2F2Q model at 310K and ff99

The thermodynamic properties of 2F2Q model of T4 Lysozyme protein vs time in (ps) during 100 ns of molecular dynamics simulation, at 310K and force field AMBER ff99, including a) Average temperature (K), b) Average pressure (bar), c) Average volume ($Å^3$), d) Average density ($amu/Å^3$), e) Average potential energy (kcal/mol), f) Average kinetic energy (kcal/mol) and g) Average total energy (kcal/mol).

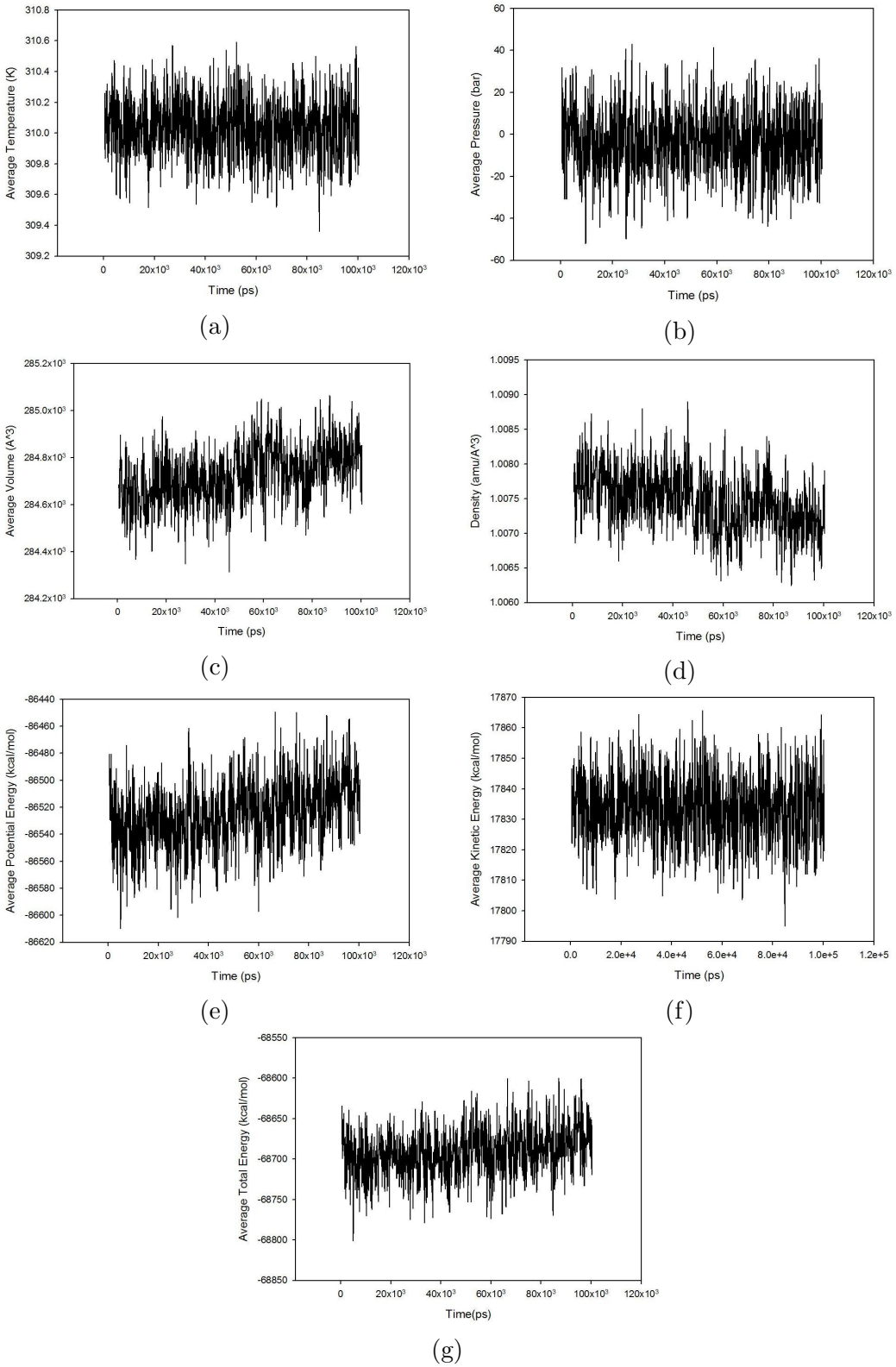


Figure 4.4: Thermodynamic properties of 2F2Q model at 310K and ff99SB

The thermodynamic properties of 2F2Q model of T4 Lysozyme protein vs time in (ps) during 100 ns of molecular dynamics simulation, at 310K and force field AMBER ff99SB, including a) Average temperature (K), b) Average pressure (bar), c) Average volume ($Å^3$), d) Average density ($amu/Å^3$), e) Average potential energy (kcal/mol), f) Average kinetic energy (kcal/mol) and g) Average total energy (kcal/mol).

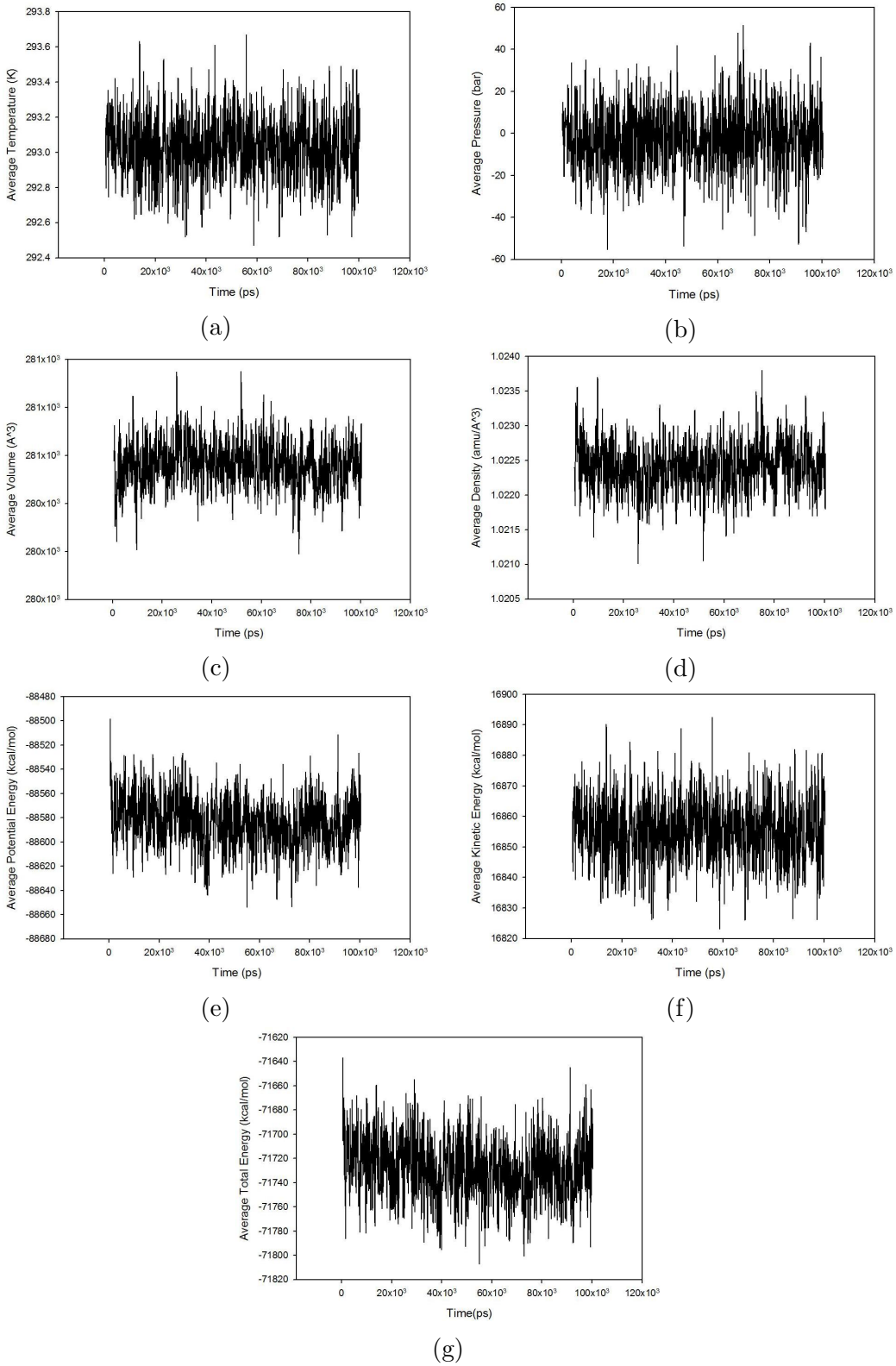


Figure 4.5: Thermodynamic properties of 2F2Q model at 293K and ff99SB

The thermodynamic properties of 2F2Q model of T4 Lysozyme protein vs time in (ps) during 100 ns of molecular dynamics simulation, at 293 K and force field AMBER ff99SB, including a) Average temperature (K), b) Average pressure (bar), c) Average volume ($Å^3$), d) Average density ($amu/Å^3$) , e) Average potential energy (kcal/mol), f) Average kinetic energy (kcal/mol) and g) Average total energy (kcal/mol).

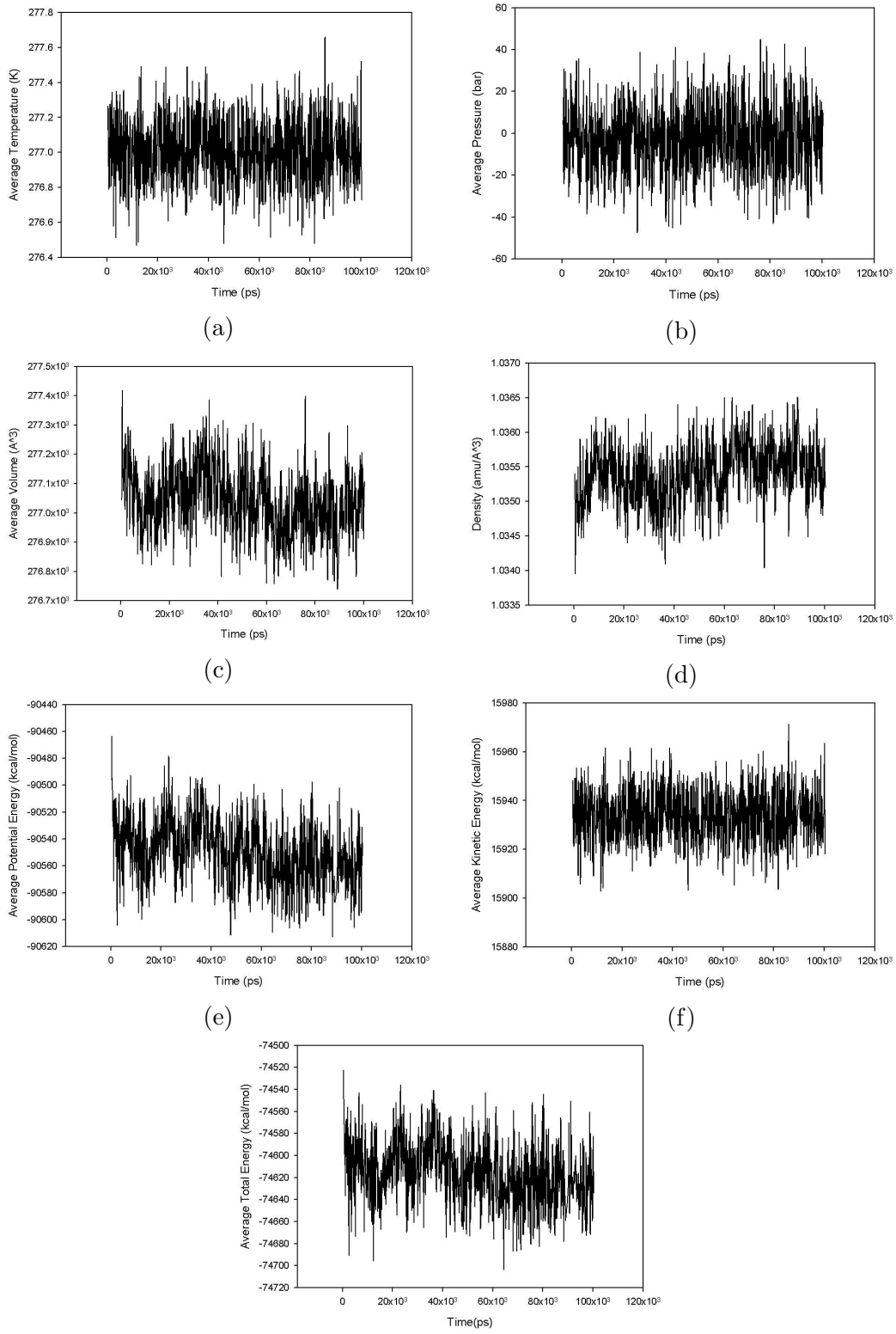


Figure 4.6: Thermodynamic properties of 2F2Q model at 277K and ff99SB

Thermodynamic properties of 2F2Q model of T4 Lysozyme protein vs time in (ps) during 100 ns of molecular dynamics simulation, at 277 K and force field AMBER ff99SB, including a) Average temperature (K), b) Average pressure (bar), c) Average volume ($\AA^3$), d) Average density ($amu/\AA^3$), e) Average potential energy (kcal/mol), f) Average kinetic energy (kcal/mol) and g) Average total energy (kcal/mol).

4.1.3 Root Mean Square Deviation (RMSD)

The root mean square deviation of a protein is a measure of the average distance between the atoms (usually the backbone atoms) with respect to a reference structure. In this case, root mean square deviation was calculated for 100,000 frames each of 1 ps of molecular dynamics simulation with respect to the initial conformation for each model, see figures,(4.7) and (4.8). For 262L model at 310K with AMBER ff99SB force field, 262L mutant model at 310K with AMBER ff99SB force field, 2F2Q model at 310K with AMBER ff99 force field, 2F2Q model at 310K with AMBER ff99SB force field, 2F2Q model at 293K with AMBER ff99SB force field and 2F2Q model at 277K with AMBER ff99SB force field, the maximum root mean square deviations were 12 $\mathring{A}$, 9 $\mathring{A}$, 16 $\mathring{A}$, 10 $\mathring{A}$ and 10 $\mathring{A}$ respectively. Early deviation occurred due to protein crystal relaxation in the surrounding water environment and equilibration. All models were generally stable during the course of simulation except for 2F2Q model at 310K with AMBER ff99 force field it became more and more weekly structured and the protein started to unfold. The root mean square deviation of the model increases rapidly with the simulation time referring to the instability of the structure, This unstable behavior was expected because of usage of AMBER ff99 force field [43,44] which is a common force field for organic and bio-organic systems. It is an older version of AMBER ff99SB force field. It was reported to cause energy imbalance between the helical and the extended regions of protein backbones and inaccurate treatment of glycine backbone parameters.[33] While AMBER ff99SB force field presents a careful reparametrization of the backbone torsion terms in ff99 and achieves much better balance of the four basic secondary structure elements.

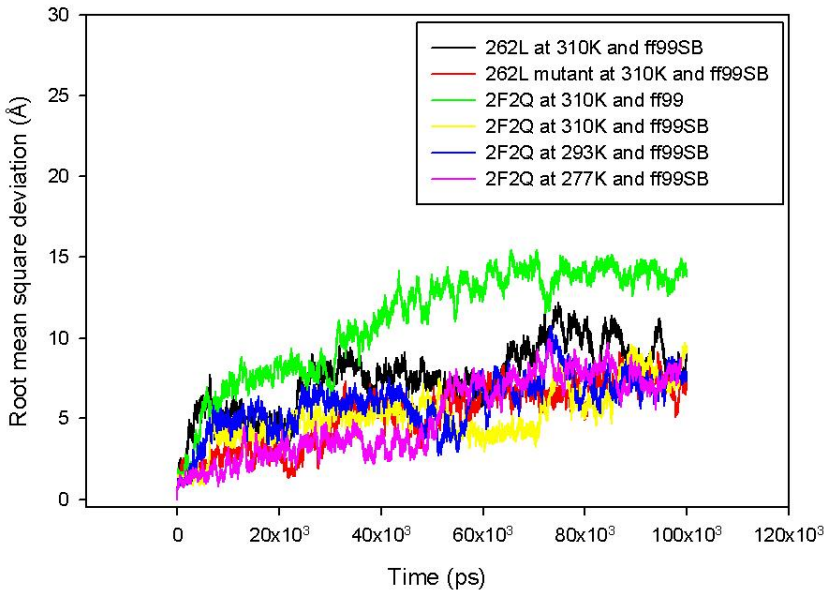


Figure 4.7: All root mean square deviations (Å) vs time (ps)

Root mean square deviation in (Å) vs time (ps) of variant T4 Lysozyme models at different conditions of temperature and force fields

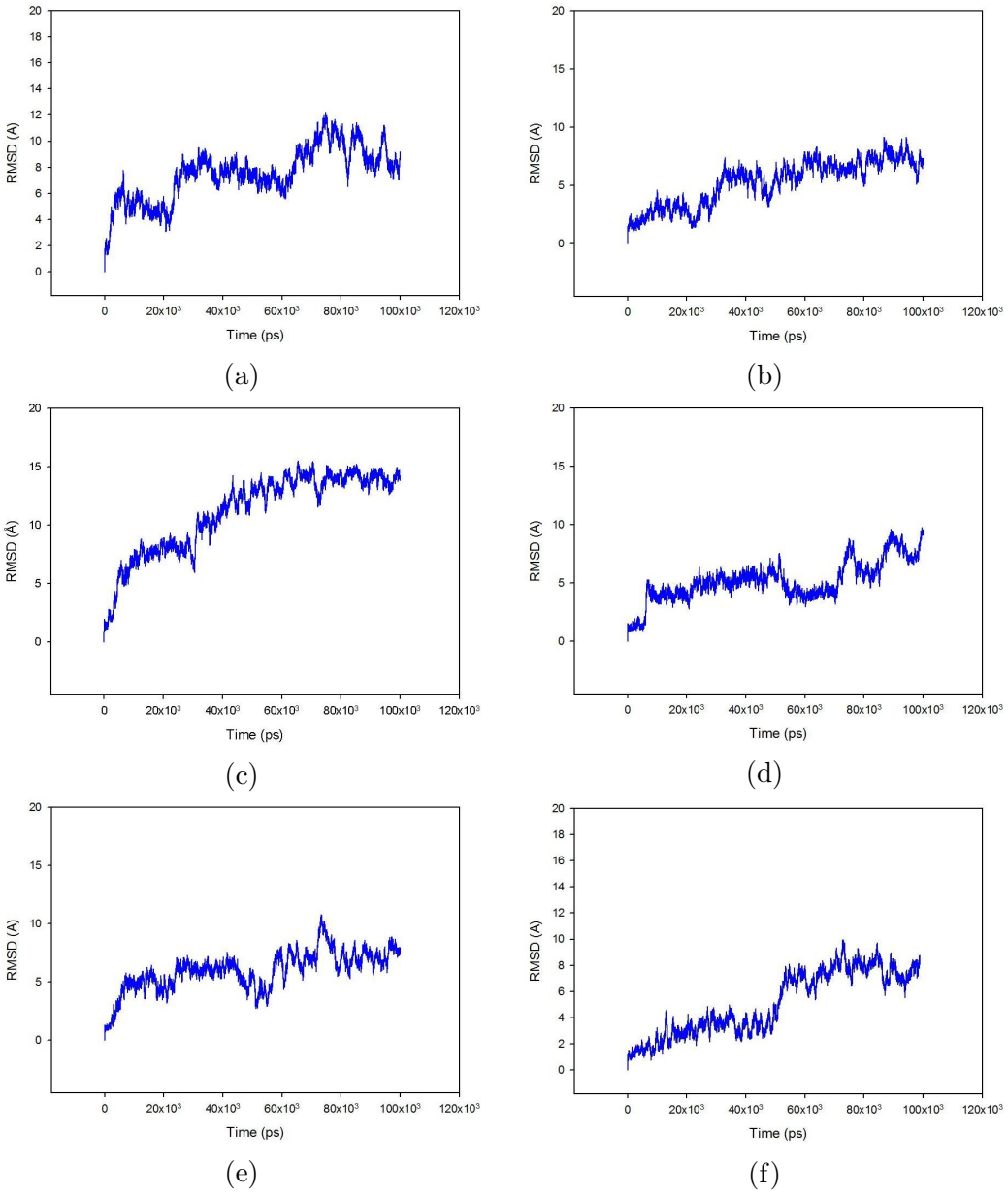


Figure 4.8: Root mean square deviation (Å) vs time (ps)

Root mean square deviation in (Å) vs time (ps) of variant T4 Lysozyme models at different conditions of temperature and force fields (a) 262L model at 310K and ff99SB force field, (b) R63A/262L mutant model at 310K and ff99SB force field, (c) 2F2Q model at 310K and ff99 force field, (d) 2F2Q model at 310K and ff99SB force field, (e) 2F2Q model at 293K and ff99SB force field and (f) 2F2Q model at 277K and ff99SB force field.

4.1.4 Dynamic Domain Analysis

For 262L mutant model at 310K with AMBER ff99SB force field, For 262L mutant model at 310K with AMBER ff99SB force field, 2F2Q model at 310K with AMBER ff99SB force field, 2F2Q model at 293K with AMBER ff99SB force field and 2F2Q model at 277K with AMBER ff99SB force field, the duplicated helix maintained its well defined structure during 100 ns of molecular dynamics simulation. The triggered motion of the duplicated helix was not detected with these models. For 2F2Q model at 310K with AMBER ff99 force field, initial stages of the triggered motion has been detected. One complete turn started to unfold from the C-terminal of the duplicated helix. Then the duplicated helix exhibited a slight shift, about $7\mathring{A}$, towards the C-terminal as shown in figure (4.9). This shift was measured from the C_α of amino acid Lysine 43 of the initial structure (t=0ns) to the same C_α of the same amino acid after 80ns of molecular dynamics simulation. Protein unfolding event occurred before detecting the complete triggered $20\mathring{A}$ towards the C-terminal. This is, again, due to usage of an old AMBER ff99 force field. It was reported to cause energy imbalance between the helical and the extended regions of protein backbones and inaccurate treatment of glycine backbone parameters.[33] More updated and advanced force field became required to keep protein stable during simulation. Time lines showing the dynamical behavior of the duplicated helix for all protein models under investigation when the β-sheets are aligned with theirs initial configurations can be found in figures (4.10), (4.11), (4.12), (4.13), (4.14) and (4.15).

Starting the molecular dynamics simulations, all models under investigation at different conditions of mutation, force fields and temperature's exhibited N-terminal hinge bending motion from its initial horizontal position of the duplicated helix as shown in figure (4.16). Characterization and analysis of these detected helical hinge motions were done using DYNDOM software. DYNDOM online software has provided an excellent tool to quantitatively analyze protein domain motions. Details of the helical hinge motions including bending simulation times, rotation angles and bending regions of amino acids in addition to the initial simulation conditions for each models can be found in table (4.1). The major domain motions of all protein models relative to their initial conformations can be found in figures (4.17), (4.18), (4.19), (4.20), (4.21) and (4.22) The perpendicular lines are crossing at the center of rotation. This fixed domain is shown in blue, the moving domain is in red and bending domain is shown in green.

Table 4.1: The hinge bending analysis by DYNDOM software.

Model No.	1	2	3	4	5	6
Model ID	262L	R63A/262L	2F2Q	2F2Q	2F2Q	2F2Q
MDS temp.	310K	310K	310K	310K	293K	277K
Force field	ff99SB	ff99SB	ff99	ff99SB	ff99SB	ff99SB
Rotation angle	90°	60°	89°	59°	89°	33.5°
Bending time	20ns	10, 30 & 60ns	20ns	50ns	40ns	40ns
Bending amino acids	(12-13) & (73-78)	(2-13) & (81-87)	(12-14) & (75-95)	(7-13) & (82-88)	(12-14) & (78-95)	(10-13) & (79-83)

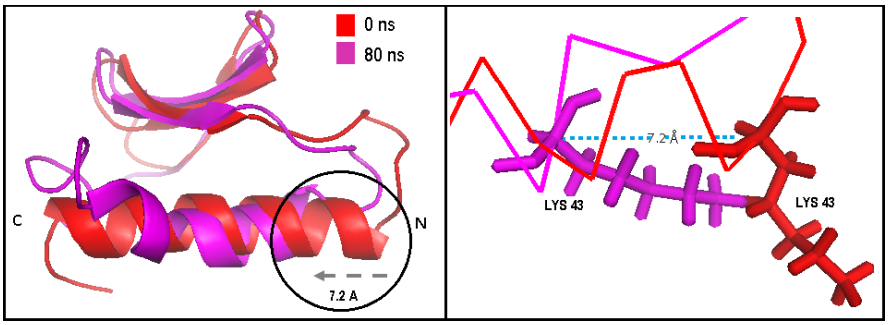


Figure 4.9: The detected triggered motion

Initial stages of the triggered motion have been detected. One turn unfolds from the C-terminal started of the duplicated helix. Then the helix exhibited a slight shift towards the C-terminal.

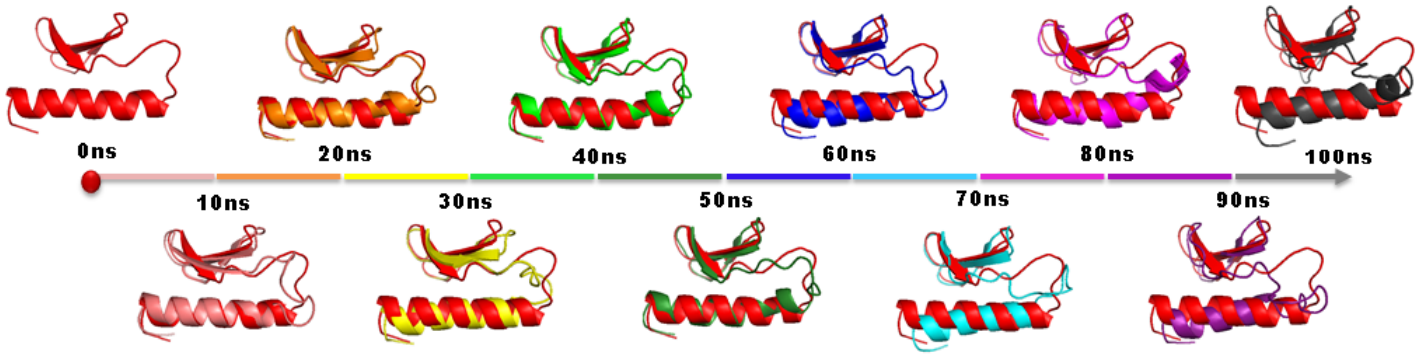


Figure 4.10: Time line of the duplicated helix of 262L model at 310K and ff99SB force field

Time line tracking the dynamical behavior of the duplicated helix compared to the initial conformation of the helix when β-sheets are aligned with theirs initial configurations of 262L model at 310K and ff99SB force field during 100ns of molecular dynamics simulation.

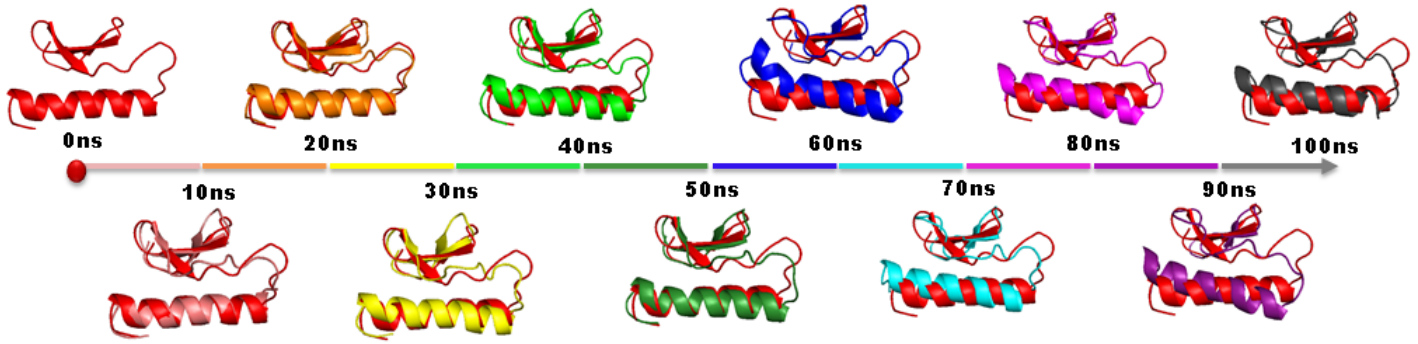


Figure 4.11: Time line of the duplicated helix of R63A/262L mutant model at 310K and ff99SB force field during 100ns of MDS

Time line tracking the dynamical behavior of the duplicated helix compared to the initial conformation of the helix when β-sheets are aligned with theirs initial configurations of R63A/262L mutant model at 310K and ff99SB force field during 100ns of molecular dynamics simulation.

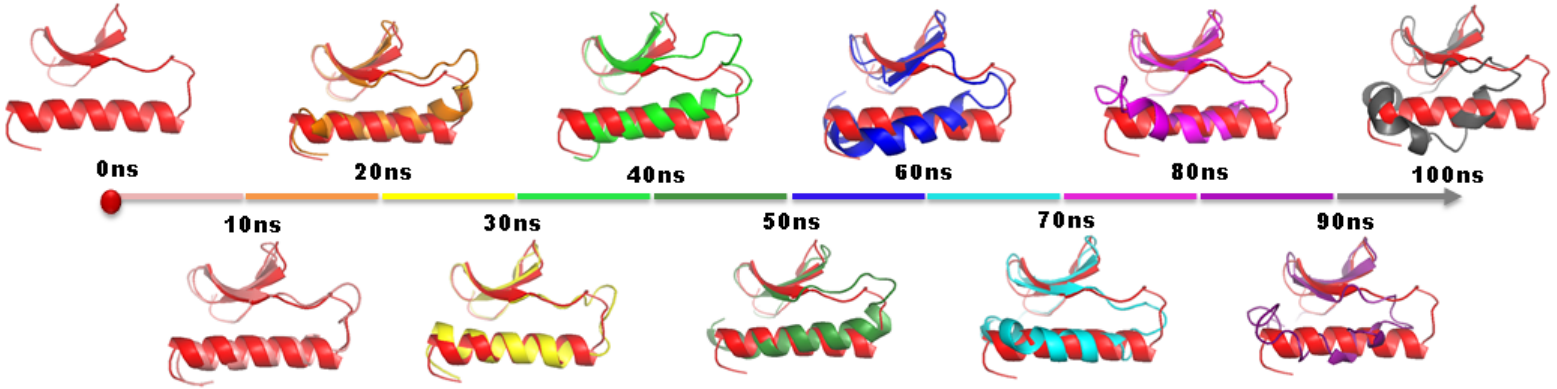


Figure 4.12: Time line of the duplicated helix of 2F2Q model at 310K and ff99 force

Time line tracking the dynamical behavior of the duplicated helix compared to the initial conformation of the helix when β-sheets are aligned with theirs initial configurations of 2F2Q model at 310K and ff99 force field during 100ns of molecular dynamics simulation.

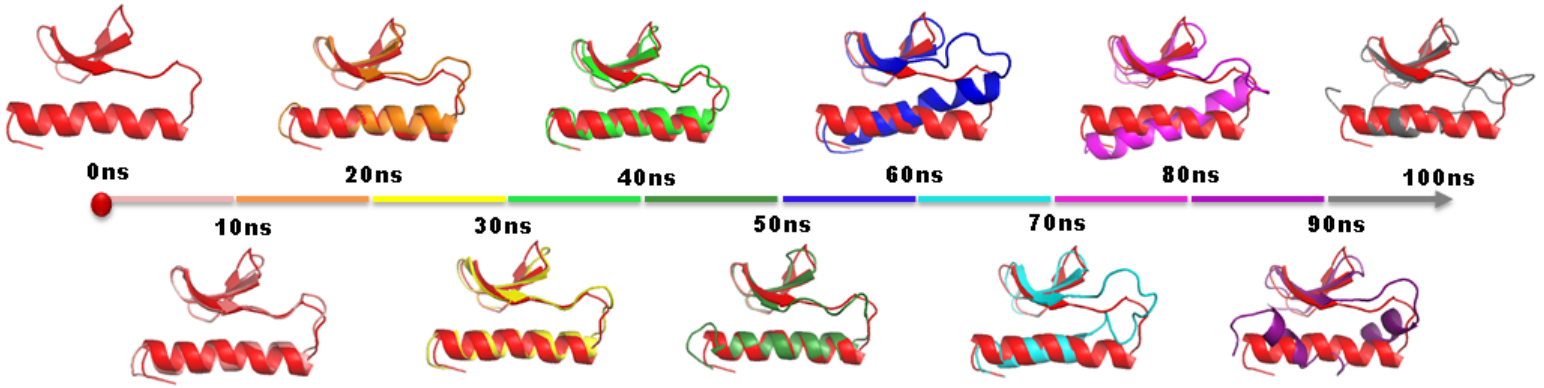


Figure 4.13: Time line of the duplicated helix of 2F2Q model at 310K and ff99SB force field

Time line tracking the dynamical behavior of the duplicated helix compared to the initial conformation of the helix when β-sheets are aligned with theirs initial configurations of 2F2Q model at 310K and ff99SB force field during 100ns of molecular dynamics simulation.

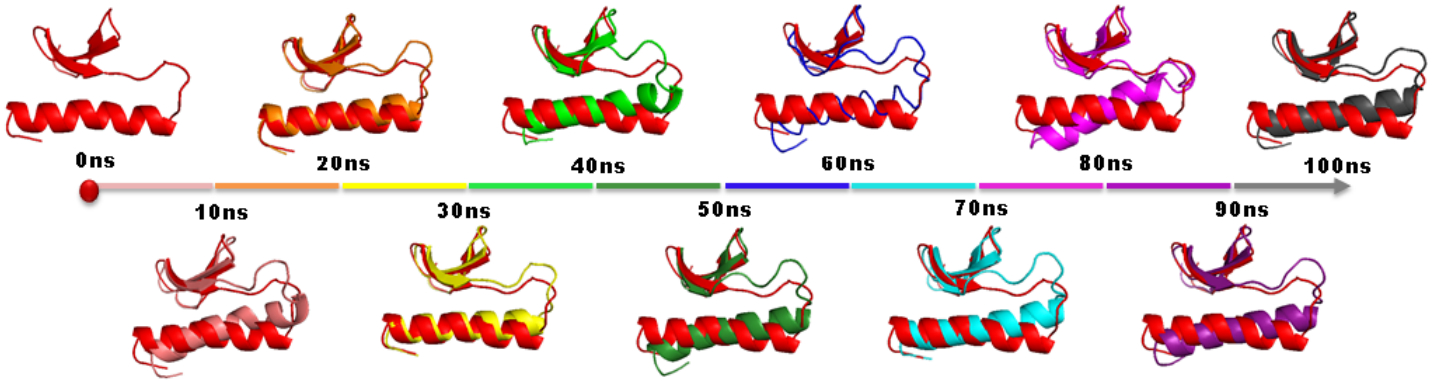


Figure 4.14: Time line of the duplicated helix of 2F2Q model at 293K and ff99SB force field

Time line tracking the dynamical behavior of the duplicated helix compared to the initial conformation of the helix when β-sheets are aligned with theirs initial configurations of 2F2Q model at 293K and ff99SB force field during 100ns of molecular dynamics simulation.

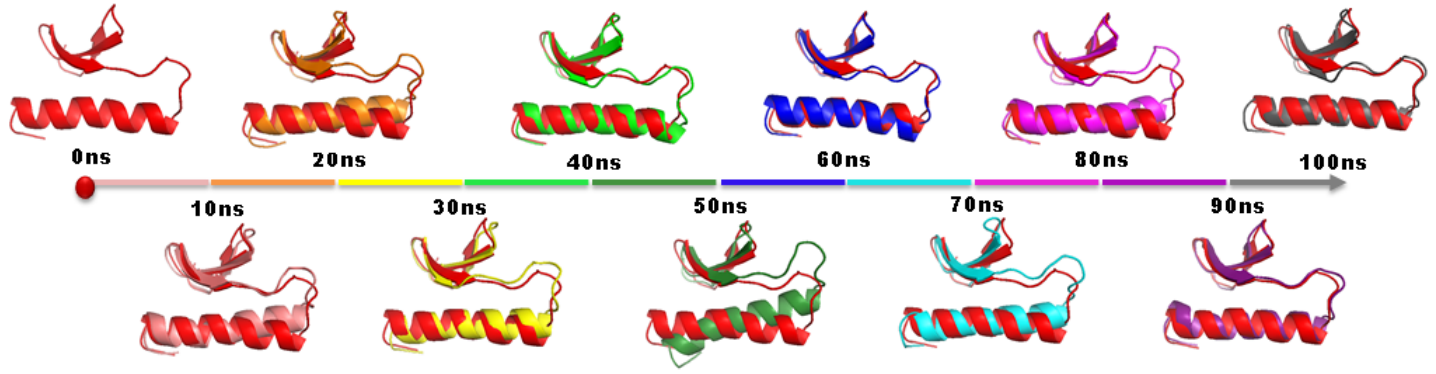


Figure 4.15: Time line of the duplicated helix of 2F2Q model at 277K and ff99SB force field

Time line tracking the dynamical behavior of the duplicated helix compared to the initial conformation of the helix when β-sheets are aligned with theirs initial configurations of 2F2Q model at 277K and ff99SB force field during 100ns of molecular dynamics simulation.

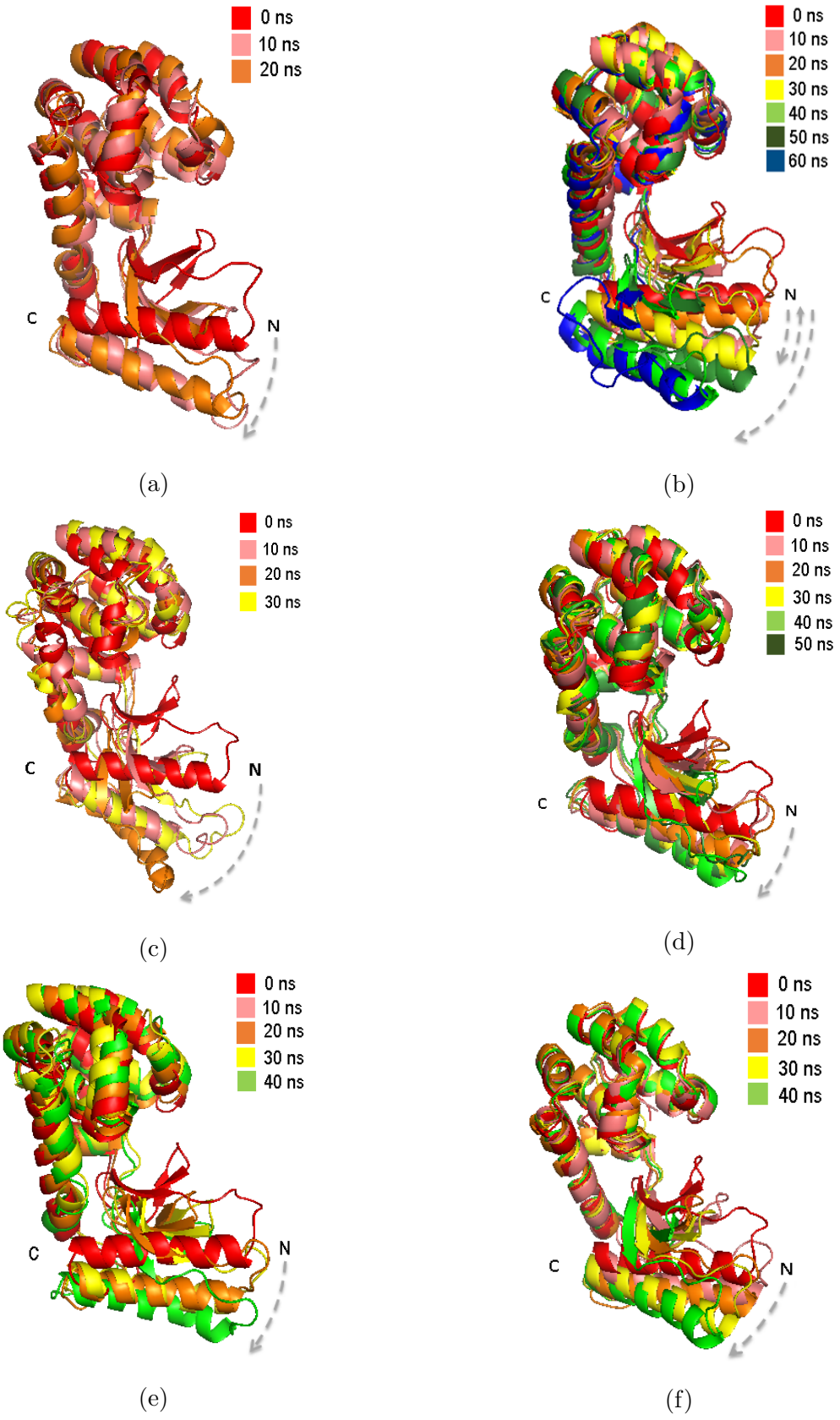


Figure 4.16: The detected hinge motion

The detected hinge motion of variant T4 Lysozyme models at different conditions of temperature and force fields (a) 262L model at 310K and ff99SB force field, (b) R63A/262L mutant model at 310K and ff99SB force field, (c) 2F2Q model at 310K and ff99 force field, (d) 2F2Q model at 310K and ff99SB force field, (e) 2F2Q model at 293K and ff99SB force field and (f) 2F2Q model at 277K and ff99SB force field.

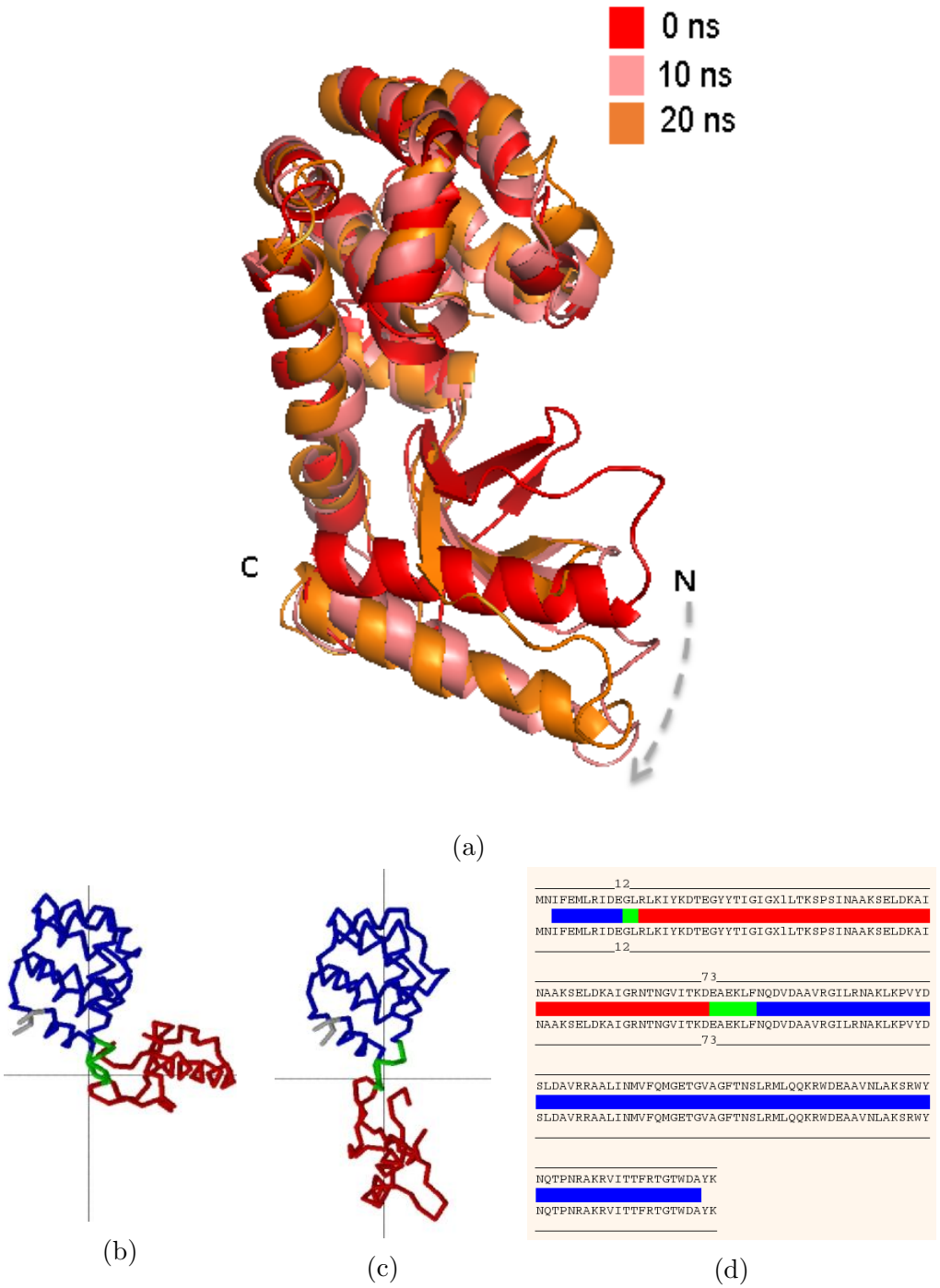


Figure 4.17: Dynamic domain analysis of 262L model at 310K and ff99SB

Major domain motion of 262L model of T4 Lysozyme protein a) Aligned snapshots over intervals of 10ns of the protein model exhibiting the hinge motion until the maximum angle of rotation , b)the initial conformation (t=0 ns), c) the conformation at (t=20 ns) and d) the corresponding protein sequence The perpendicular lines are crossing at the center of rotation. This fixed domain is shown in blue, the moving domain is in red and bending domain is shown in green. The N-terminal of the duplicated helix exhibited a hinge bending motion with a rotation angle of 90 degrees. The bending that lead to the hinge bending motion happened around (12-13) and (73-78) amino acids.

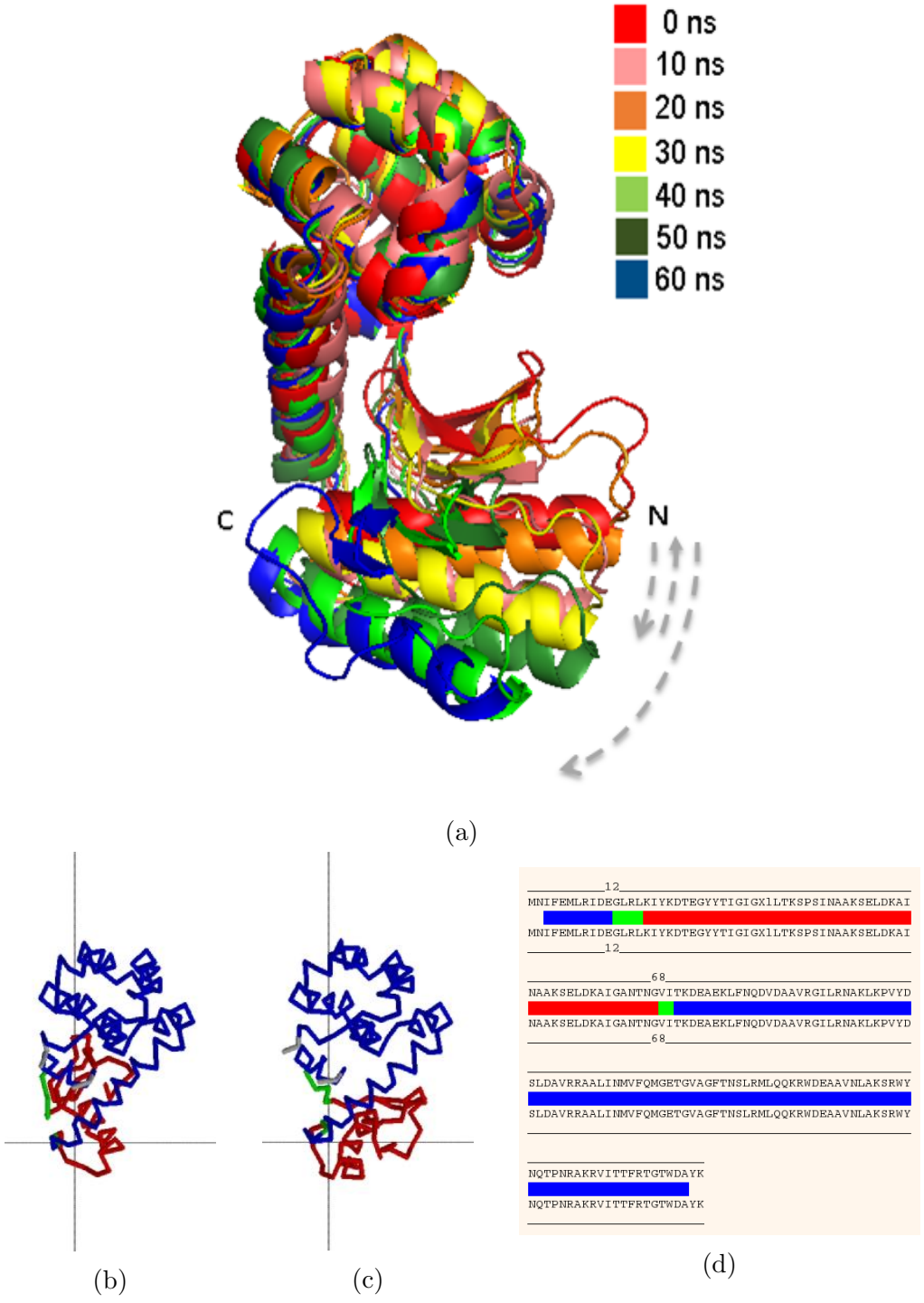


Figure 4.18: Dynamic Domain analysis of 262L mutant model at 310K and ff99SB

Major domain motion of 262L mutant model of T4 Lysozyme protein relative to the initial conformation including a) aligned snapshots over intervals of 10ns of the protein model during the simulation exhibiting the hinge motion until the maximum angle of rotation, b) the initial conformation (t=0 ns), c) the conformation at (t=30 ns) and d) the corresponding protein sequence. The perpendicular lines are crossing at the center of rotation. This fixed domain is shown in blue, the moving domain is in red and bending domain is shown in green. The N-terminal of the duplicate helix exhibited a hinge bending motion with a rotation angle of 60 degrees. The bending that lead to the hinge bending motion happened around (3-13) and (81-87) amino acids.

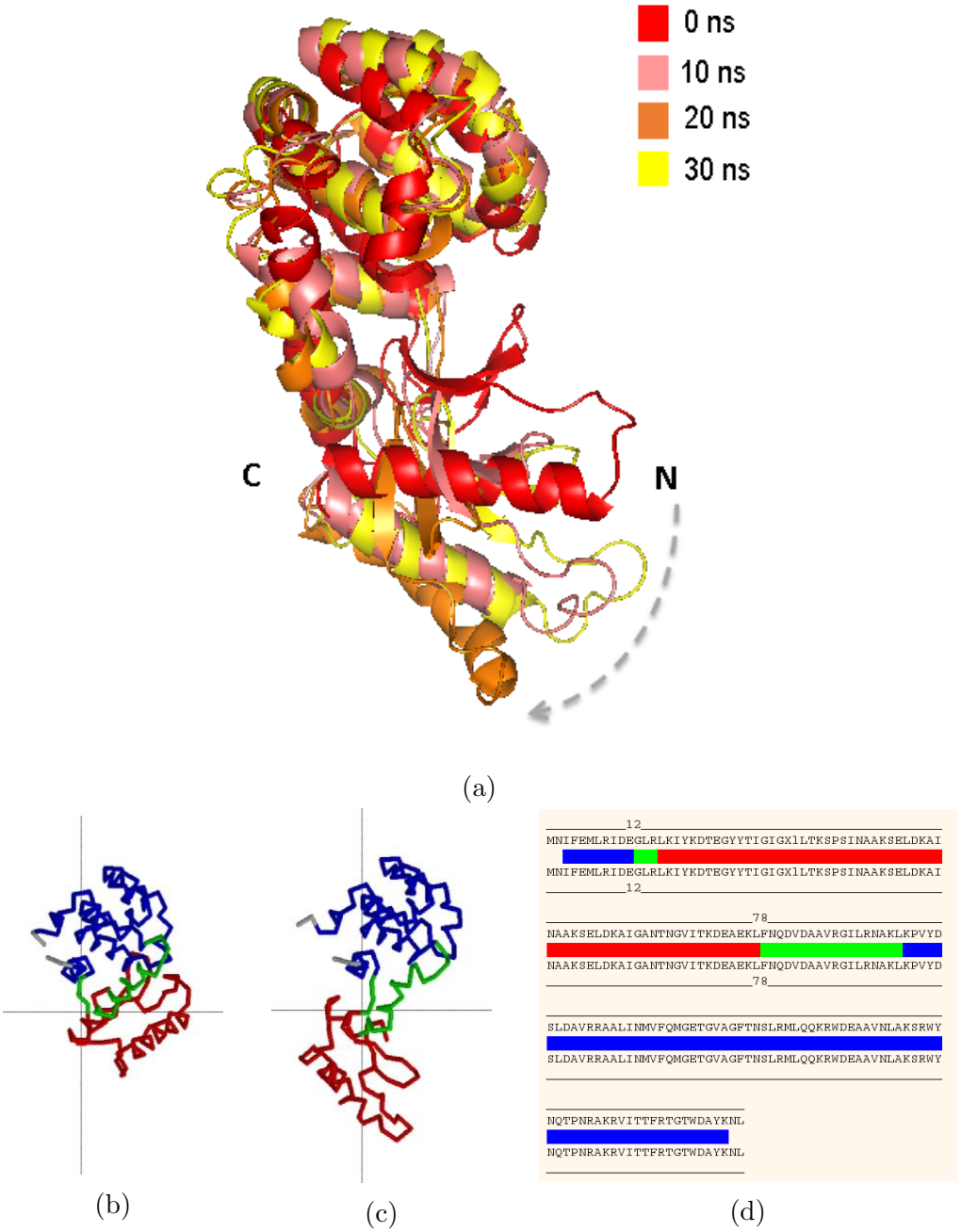


Figure 4.19: Dynamic domain analysis of 2F2Q model at 310K and ff99

Major domain motion of 2F2Q model of T4 Lysozyme protein, at 310K and force field AMBER ff99, relative to the initial conformation including a) aligned snapshots over intervals of 10ns of the protein model during the simulation exhibiting the hinge motion until the maximum angle of rotation, b) the initial conformation (t=0 ns), c) the conformation at (t=20 ns) and d) the corresponding protein sequence. The perpendicular lines are crossing at the center of rotation. This fixed domain is shown in blue, the moving domain is in red and bending domain is shown in green. The N-terminal of the duplicate helix exhibited a hinge bending motion with a rotation angle of 89 degrees. The bending that lead to the hinge bending motion happened around (12-14) and (78-95) amino acids.

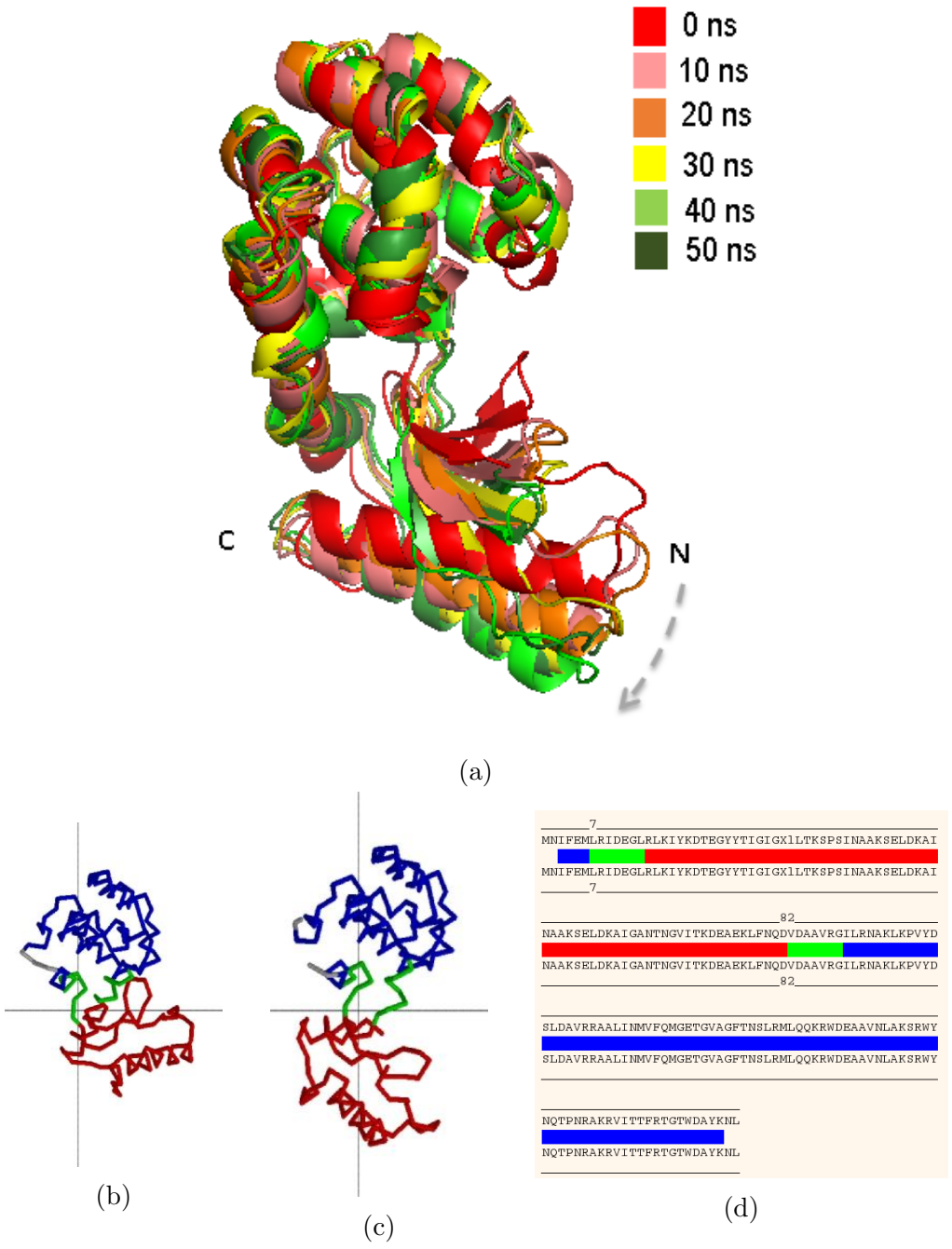


Figure 4.20: Dynamic domain analysis of 2F2Q model at 310K and ff99SB

Major domain motion of 2F2Q model of T4 Lysozyme protein, at 310K and force field AMBER ff99SB, relative to the initial conformation including a) aligned snapshots over intervals of 10ns of the protein model during the simulation exhibiting the hinge motion until the maximum angle of rotation, b) the initial conformation (t=0 ns), c) the conformation at (t=50 ns) and d) the corresponding protein sequence. The perpendicular lines are crossing at the center of rotation. This fixed domain is shown in blue, the moving domain is in red and bending domain is shown in green. The N-terminal of the duplicate helix exhibited a hinge bending motion with a rotation angle of 59 degrees. The bending that lead to the hinge bending motion happened around (7-13) and (82-88) amino acids.

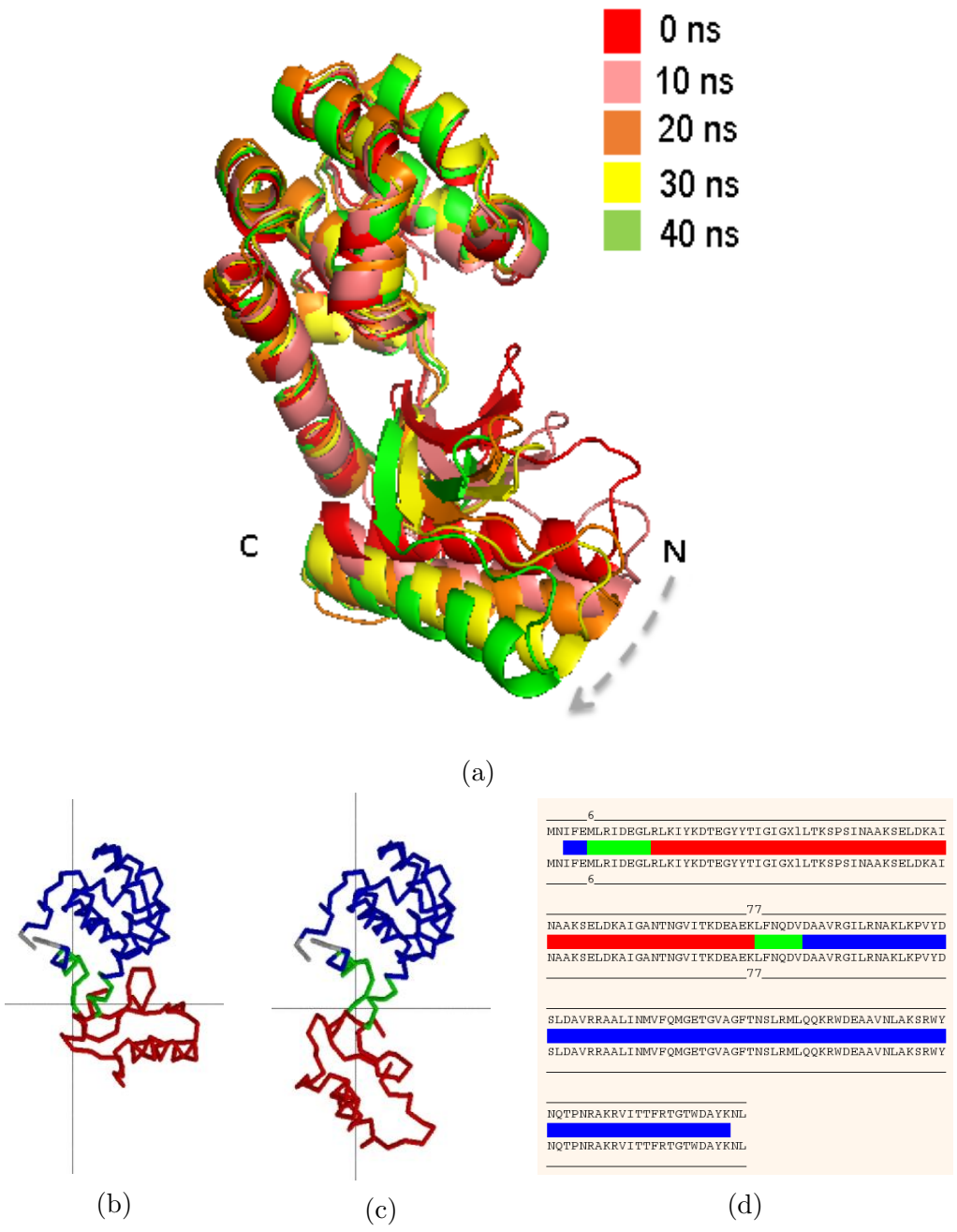


Figure 4.21: Dynamic Domain analysis of 2F2Q model at 293K and ff99SB

Major domain motion of 2F2Q model of T4 Lysozyme protein, at 293 K and force field AMBER ff99SB, relative to the initial conformation including a) aligned snapshots over intervals of 10ns of the protein model during the simulation exhibiting the hinge motion until the maximum angle of rotation, b) the initial conformation (t=0 ns), c) the conformation at (t=40 ns) and d) the corresponding protein sequence. The perpendicular lines are crossing at the center of rotation. This fixed domain is shown in blue, the moving domain is in red and bending domain is shown in green. The N-terminal of the duplicate helix exhibited a hinge bending motion with a rotation angle of 89 degrees. The bending that lead to the hinge bending motion happened around (12-14) and (78-95) amino acids.

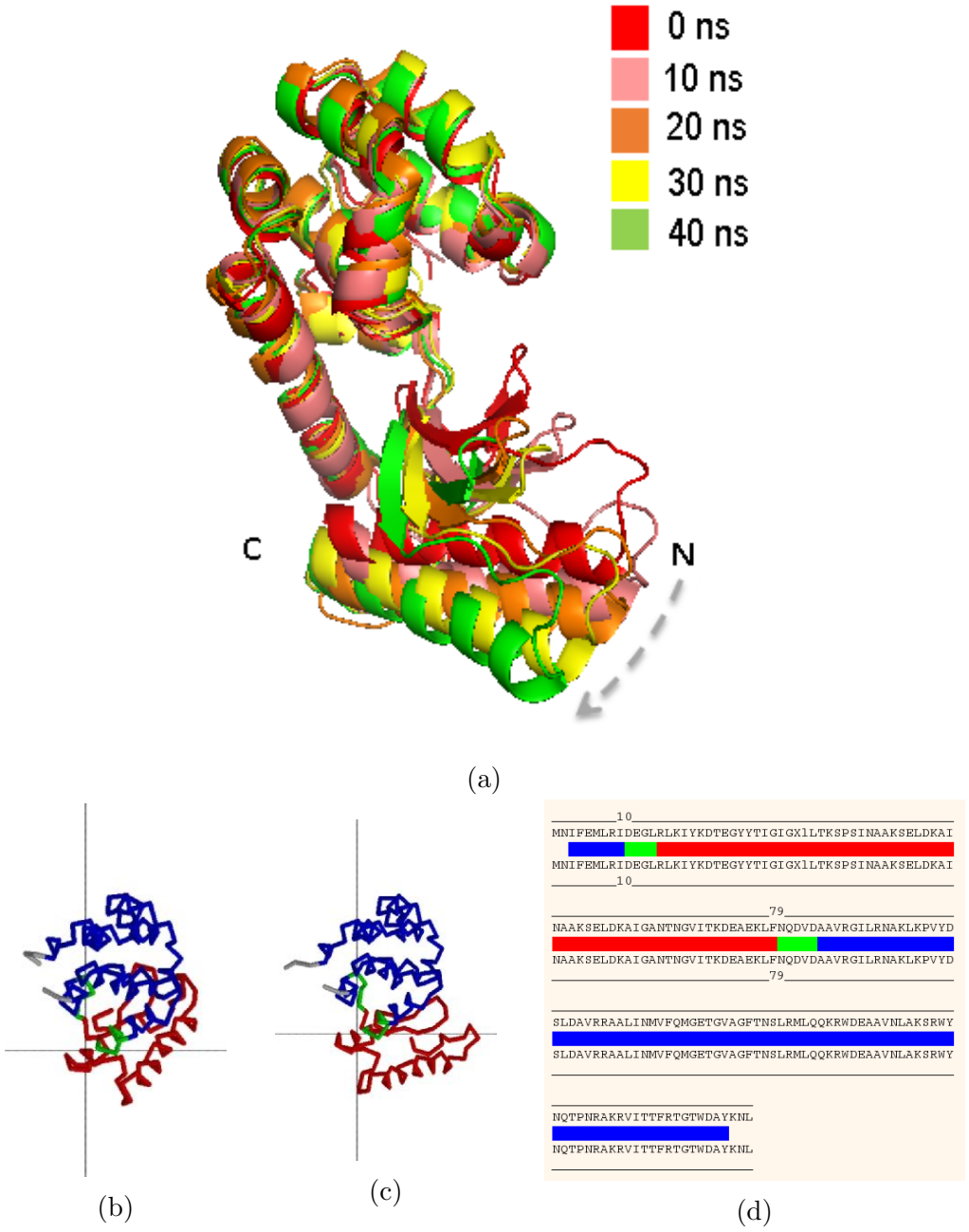


Figure 4.22: Dynamic domain analysis of of 2F2Q model at 277K and ff99SB

Major domain motion of 2F2Q model of T4 Lysozyme protein, at 277 K and force field AMBER ff99SB, relative to the initial conformation including a) aligned snapshots over intervals of 10ns of the protein model during the simulation exhibiting the hinge motion until the maximum angle of rotation, b) the initial conformation (t=0 ns), c) the conformation at (t=40 ns) and d) the corresponding protein sequence. The perpendicular lines are crossing at the center of rotation. This fixed domain is shown in blue, the moving domain is in red and bending domain is shown in green. The N-terminal of the duplicate helix exhibited a hinge bending motion with a rotation angle of 33.5 degrees. The bending that lead to the hinge bending motion happened around (10-13) and (79-83) amino acids.

4.1.5 Hydrogen Bonding Analysis of Amino acid 63

Hydrogen bonds have played an incredibly important role in protein stability and structure. They provide the major stabilization for the regular secondary structure for proteins namely α-helices and β-sheets. 'The energy of a hydrogen bond is about 5 kcal/mol and this is the same order of magnitude as the gibbs energy for protein denaturation which means that leaving one hydrogen bond unmade could amount to denaturating a protein. In fact, protein structure show that all possible hydrogen bonds are stabilized in native state either within the protein or with water. A hydrogen bond X-H...Y is an attractive interaction between a group X-H in which a covalent bond exists between electronegative atom X and the H atom, and another electronegative atom or group. The electropositive H atom inserts between the two electronegative atoms X and Y bringing them together the X-H group is the hydrogen donor and the Y group is the hydrogen acceptor. The X and Y atoms are typically oxygen, nitrogen and fluorine. A hydrogen bond has a large electrostatic component, which consists of the interaction between the positively polarized end of the X-H dipole (the H atom) and the negatively polarized end of the dipole ending at the electronegative dipole atom. There is also a covalent component arising from charge transfer between donor and acceptor and a dispersion component. [50]

For 262L model at 310K and ff99SB force field, 28 different hydrogen bonds with the atoms of Arginine 63 have been detected providing a cut off distance (acceptor to donor heavy atom) of 3.0 Å. 23 Hydrogen bonds (1-23) are formed with atoms of guanidinium head group of Arginine 63 only 6 hydrogen bonds of them (1, 4, 8, 9, 15 19) were resident for more than 5% of the simulation as shown in figures (4.23) and (4.24a). 4 Hydrogen bonds (24-28) are formed with the backbone of Arginine 63 and no one of them was resident for more than 5% of the simulation.

For 262L mutant model at 310K and ff99SB force field, 6 different hydrogen bonds with the atoms of Alanine 63 have been detected providing a cut off distance (acceptor to donor heavy atom) of 3.0 Å. For 2F2Q model at 310K and ff99 force field, 15 different hydrogen bonds with the atoms of Alanine 63 have been detected. For 2F2Q model at 310K and ff99SB force field, 9 different hydrogen bonds with the atoms of Alanine 63 have been detected. For 2F2Q model at 293K and ff99SB force field and 2F2Q model at 277K and ff99SB force field, 4 different hydrogen bonds with the atoms of Alanine 63 have been detected . All hydrogen bonds detected for 262L mutant model at 310K and ff99SB force field, 2F2Q model at 310K and ff99

force field, 2F2Q model at 310K and ff99SB force field, 2F2Q model at 293K and ff99SB force field and 2F2Q model at 277K and ff99SB force field are formed with the backbone of Alanine 63 providing a cut off distance (acceptor to donor heavy atom) of 3.0 Åand no one of them was resident for more than 5% of the simulation as shown in figures(4.24b), (4.24c), (4.24d), (4.24e) and (4.24f) respectively.

Details of hydrogen bonds involving amino acid Amino acid 63 including donor atoms, hydrogen atoms, acceptor atoms, the fraction of frames the bond is present, bond length, bond angle and the presence percentage of the hydrogen bonds during 100ns of simulation for 262L model at 310K and ff99SB force field, R63A/262L mutant model at 310K and ff99SB force field, 2F2Q model at 310K and ff99 force field, 2F2Q model at 310K and ff99SB force field, 2F2Q model at 293K and ff99SB force field and 2F2Q model at 277K and ff99SB force field respectively can be found in tables (4.2), (4.3), (4.4), (4.5), (4.6), (4.7).

Table 4.2: Hydrogen bonds of Arginine 63 of 262l model at 310K and ff99SB

Index	Acceptor atom	Hydrogen atom	Donor atom	Average length (Å)	Average angle (°)	Percentage
1	ILE 61/O	ARG 63/HH11	ARG 63/NH1	2.9 ±0.3	158 ±32	14.7
2	ASN 64/O	ARG 63/HH11	ARG 63/NH1	2.8 ±0.2	150 ±25	1.4
3	GLY 62/O	ARG 63/HH11	ARG 63/NH1	2.9 ±0.2	158 ±23	1.1
4	GLU 73/OE2	ARG 63/HH12	ARG 63/NH1	2.8 ±0.3	163 ±47	38.9
5	GLU 73/OE1	ARG 63/HH12	ARG 63/NH1	2.8 ±0.2	159 ±33	2.6
6	ASP 72/OD2	ARG 63/HH12	ARG 63/NH1	2.8 ±0.2	158 ±28	0.0
7	THR 70/OG1	ARG 63/HH12	ARG 63/NH1	2.9 ±0.2	154 ±28	0.0
8	GLU 73/OE1	ARG 63/HH21	ARG 63/NH2	2.8 ±0.3	158 ±48	14.3
9	GLU 73/OE2	ARG 63/HH21	ARG 63/NH2	2.8 ±0.2	150 ±48	4.1
10	ASN 64/O	ARG 63/HH21	ARG 63/NH2	2.8 ±0.2	149 ±54	0.3
11	THR 70/OG1	ARG 63/HH21	ARG 63/NH2	2.9 ±0.3	149 ±31	0.0
12	ILE 61/O	ARG 63/HH21	ARG 63/NH2	2.8 ±0.3	148 ±32	0.0
13	ASP 72/OD1	ARG 63/HH21	ARG 63/NH2	3.0 ±0.3	140 ±37	0.0
14	GLU 73/OE1	ARG 63/HH22	ARG 63/NH2	2.8 ±0.3	163 ±35	36.3
15	GLU 73/OE2	ARG 63/HH22	ARG 63/NH2	2.8 ±0.3	154 ±43	6.3
16	THR 70/OG1	ARG 63/HH22	ARG 63/NH2	2.9 ±0.3	151 ±32	0.1
17	VAL 68/O	ARG 63/HH22	ARG 63/NH2	2.9 ±0.3	151 ±27	0.0
18	ASP 72/OD2	ARG 63/HH22	ARG 63/NH2	2.9 ±0.3	144 ±31	0.0
19	GLU 73/OE2	ARG 63/HE	ARG 63/NE	2.8 ±0.3	158 ±46	8.9
20	GLU 73/OE1	ARG 63/HE	ARG 63/NE	2.9 ±0.3	150 ±36	2.7
21	ASN 64/O	ARG 63/HE	ARG 63/NE	2.9 ±0.3	154 ±36	0.3
22	ILE 61/O	ARG 63/HE	ARG 63/NE	2.9 ±0.3	156 ±48	0.2
23	GLY 62/O	ARG 63/HE	ARG 63/NE	2.9 ±0.3	161 ±25	0.0
24	ILE 61/O	ARG 63/H	ARG 63/N	2.9 ±0.6	144 ±16	2.2
25	ASP 58/O	ARG 63/H	ARG 63/N	2.9 ±0.6	143 ±19	0.2
26	ASP 58/OD2	ARG 63/H	ARG 63/N	2.9 ±0.6	138 ±17	0.0
27	ARG 63/O	THR 65/HG1	THR 65/OG1	2.9 ±0.6	161 ±24	0.4
28	ARG 63/O	THR 65/H	THR 65/N	2.9 ±0.6	143 ±7	0.3

Table 4.3: Hydrogen bonds of Alanine 63 of 262L mutant model at 310K and ff99SB

Index	Acceptor atom	Hydrogen atom	Donor atom	Average length (Å)	Average angle (°)	Percentage
1	ILE 61/O	ALA 63/H	ALA 63/N	2.9 ±0.6	146 ±20	1.8
2	ASP 58/O	ALA 63/H	ALA 63/N	2.9 ±0.6	147 ±14	0.5
3	LEU 57/O	ALA 63/H	ALA 63/N	2.9 ±0.6	152 ±33	0.5
4	ASP 58/OD2	ALA 63/H	ALA 63/N	2.8 ±0.4	148 ±24	0.0
5	ALA 63/O	THR 65/H	THR 65/N	2.9 ±0.4	145 ±28	3.6
6	ALA 63/O	THR 65/HG1	THR 65/OG1	2.8 ±0.6	161 ±24	1.1

Table 4.4: Hydrogen bonds of Alanine 63 of 2F2Q model at 310K and ff99

Index	Acceptor atom	Hydrogen atom	Donor atom	Average length (Å)	Average angle (°)	Percentage
1	ALA 60/O	ALA 63/H	ALA 63/N	2.9 ±0.2	158 ±23	9.4
2	LYS 59/O	ALA 63/H	ALA 63/N	2.9 ±0.2	158 ±25	1.6
3	ASP 58/O	ALA 63/H	ALA 63/N	2.9 ±0.3	156 ±24	0.9
4	ILE 61/O	ALA 63/H	ALA 63/N	2.9 ±0.2	146 ±19	0.7
5	ASP 58/OD2	ALA 63/H	ALA 63/N	2.9 ±0.3	144 ±29	0.0
6	ASP 58/OD1	ALA 63/H	ALA 63/N	2.9 ±0.3	146 ±28	0.0
7	ALA 63/O	ASN 66/H	ASN 66/N	2.9 ±0.3	158 ±21	2.9
8	ALA 63/O	GLY 67/H	GLY 67/N	2.9 ±0.2	150 ±44	2.9
9	ALA 63/O	THR 65/HG1	THR 65/OG1	2.9 ±0.3	164 ±28	1.1
10	ALA 63/O	THR 65/H	THR 65/N	2.9 ±0.5	146 ±17	0.8
11	ALA 63/O	LEU 15/H	LEU 15/N	2.9 ±0.2	158 ±46	0.5
12	ALA 63/O	VAL 68/H	VAL 68/N	2.9 ±0.2	153 ±30	0.4
13	ALA 63/O	ASN 64/HD22	ASN 64/ND2	2.9 ±0.2	153 ±19	0.0
14	ALA 63/O	TYR 18/HH	TYR 18/OH	2.9 ±0.2	155 ±50	0.0
15	ALA 63/O	ASN 66/HD22	ASN 66/ND2	2.9 ±0.2	153 ±23	0.0

Table 4.5: Hydrogen bonds of Alanine 63 of 2F2Q model at 310K and ff99SB

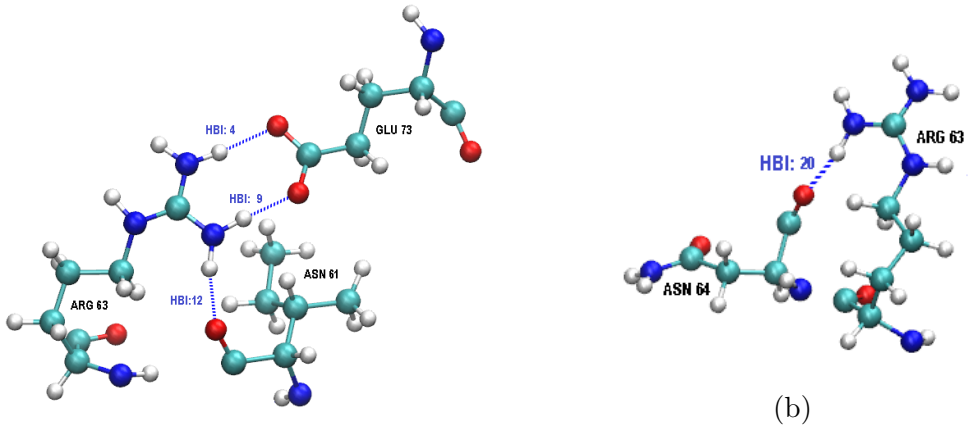
Index	Acceptor atom	Hydrogen atom	Donor atom	Average length (Å)	Average angle (°)	Percentage
1	ILE 61/O	ALA 63/H	ALA 63/N	2.9 ±0.7	146 ±21	7.5
2	ASP 58/OD1	ALA 63/H	ALA 63/N	2.8 ±0.8	155 ±25	0.2
3	ASP 58/O	ALA 63/H	ALA 63/N	2.9 ±0.6	145 ±27	0.1
4	ASP 58/OD2	ALA 63/H	ALA 63/N	2.8 ±0.6	154 ±25	0.1
5	ALA 63/O	THR 65/H	THR 65/N	2.9 ±0.5	145 ±18	1.5
6	ALA 63/O	THR 65/HG1	THR 65/OG1	2.9 ±0.8	164 ±21	1.1
7	ALA 63/O	LYS 76/HZ3	LYS 76/NZ	2.9 ±0.3	159 ±27	0.1
8	ALA 63/O	LYS 76/HZ2	LYS 76/NZ	2.9 ±0.3	158 ±28	0.1
9	ALA 63/O	LYS 76/HZ1	LYS 76/NZ	2.9 ±0.3	157 ±25	0.1

Table 4.6: Hydrogen bonds of Alanine 63 of 2F2Q model at 293K and ff99SB

Index	Acceptor atom	Hydrogen atom	Donor atom	Average length (Å)	Average angle (°)	Percentage
1	ILE 61/O	ALA 63/H	ALA 63/N	2.9 ±0.6	146 ±16	2.1
2	ASP 58/O	ALA 63/H	ALA 63/N	2.9 ±0.4	146 ±20	1.4
3	ASP 58/OD1	ALA 63/H	ALA 63/N	2.9 ±0.4	158 ±20	0.1
4	ALA 63/O	THR 65/H	THR 65/N	2.9 ±0.6	145 ±20	3.2

Table 4.7: Hydrogen bonds of Alanine 63 of 2F2Q model at 277K and ff99SB

Index	Acceptor atom	Hydrogen atom	Donor atom	Average length (Å)	Average angle(°)	Percentage
1	ILE 61/O	ALA 63/H	ALA 63/N	2.9 ±0.8	146 ±17	3.1
2	ASP 58/O	ALA 63/H	ALA 63/N	2.9 ±0.7	145 ±21	0.7
3	ASP 58/OD2	ALA 63/H	ALA 63/N	2.8 ±0.4	141 ±21	0.0
4	ALA 63/O	THR 65/H	THR 65/N	2.9 ±0.4	146 ±17	1.2



(a) (b)

Figure 4.23: The dominant hydrogen bonds

Illustration of some dominant hydrogen bonds with the guanidinium head group of Arginine 63 of 262L model at 310K and ff99SB. Information of illustrated hydrogen bonds can be found in table (4.2) with a) indices 4, 9 and 12 b)index 20.

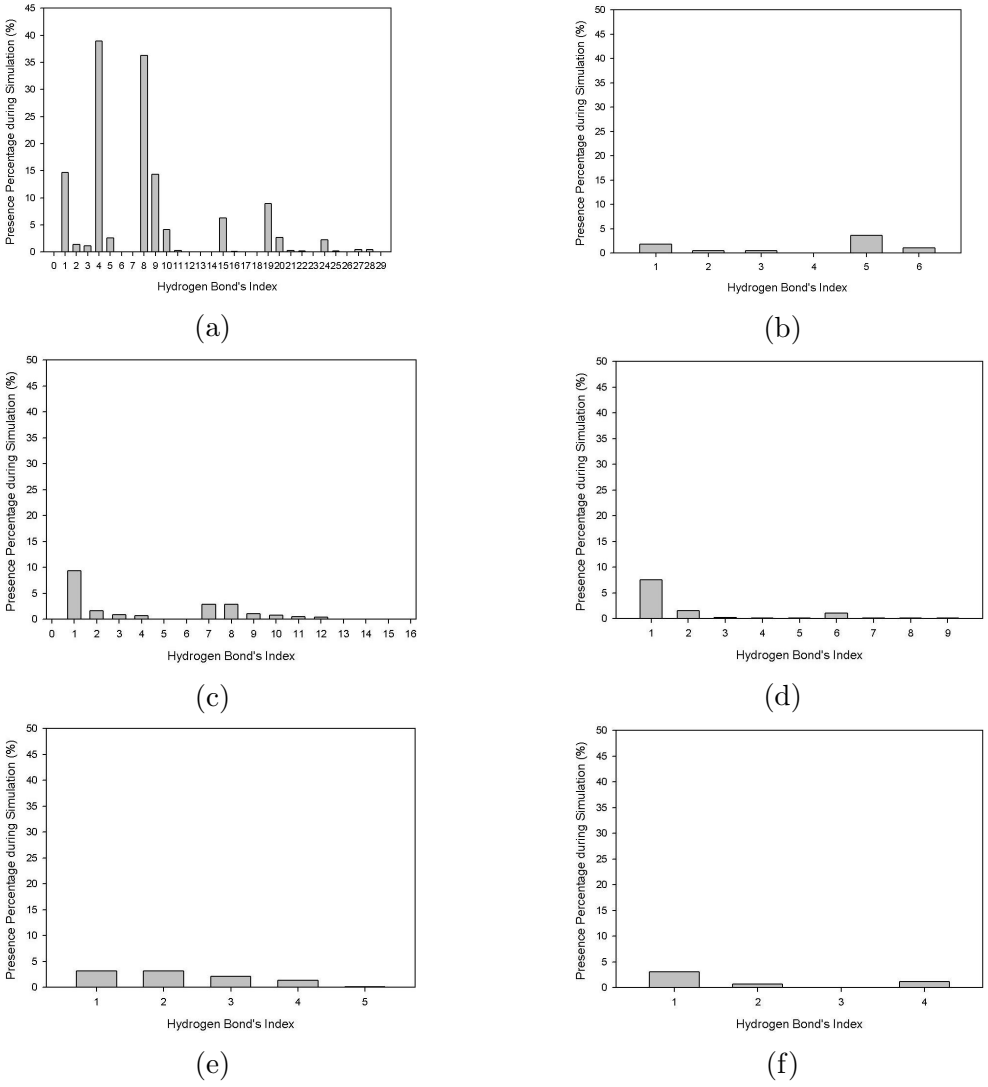


Figure 4.24: Presence percentage of hydrogen bonds in Amino acid 63

Presence percentage of hydrogen bonds in Amino acid 63 of variant T4 Lysozyme models during 100 ns of molecular dynamics simulation at different conditions of temperature and force field (a) 262L model at 310K and ff99SB force field, (b) R63A/262L mutant model at 310K and ff99SB force field, (c) 2F2Q model at 310K and ff99 force field, (d) 2F2Q model at 310K and ff99SB force field, (e) 2F2Q model at 293K and ff99SB force field and (f) 2F2Q model at 277K and ff99SB force field.

4.1.6 Radial Distribution Function (RDF)

The average relative location of the atoms of a liquid are expressed in terms of the radial distribution function, $g(r)$. This function is defined so that $g(r)r^2dr$ is the probability that an atom will be found in the range dr at a distance r from another atom. Because all the mentioned hydrogen bonds of amino acid 63 have been detected providing a cut off distance (acceptor to donor heavy atom) of 3.0Å, calculating the radial distribution function (RDF) gave us information about the actual average length and its probability of the detected hydrogen bonds even if it is further longer than 3.0Å. For 262L model at 310K with AMBER ff99SB force field, radial distribution functions around donors and acceptors atoms in Arginine 63 including a)ARG 63/NH1, b)ARG 63/NH2, c)ARG 63/NHE, d)ARG 63/N and e)ARG 63/O with all corresponding acceptors and donors atoms from any other amino acid for all detected hydrogen bonds in table (4.2) are shown in figure (4.25). For 262L mutant model at 310K and ff99SB force field, 2F2Q model at 310K and ff99 force field, 2F2Q model at 310K and ff99SB force field, 2F2Q model at 293K and ff99SB force field and 2F2Q model at 277K and ff99SB force field, radial distribution functions around donors and acceptors atoms in Alanine 63 including a)ARG 63/N and b)ARG 63/O with all corresponding acceptors and donors atoms from any other amino acid for all detected hydrogen bonds in tables (4.3), (4.4), (4.5), (4.6) and (4.7) are shown in figures (4.26), (4.27), (4.28), (4.29) and (4.30) respectively. It is worth mentioning that for all protein models where the updated ff99SB is the used force field, the plots representing the radial distribution function were smooth. While for the protein model where the ff99 is the used force field, all the plots representing the radial distribution functions between atoms of all the detected hydrogen bonds were rough and meandrous. This may be an artifact of this very old force field.

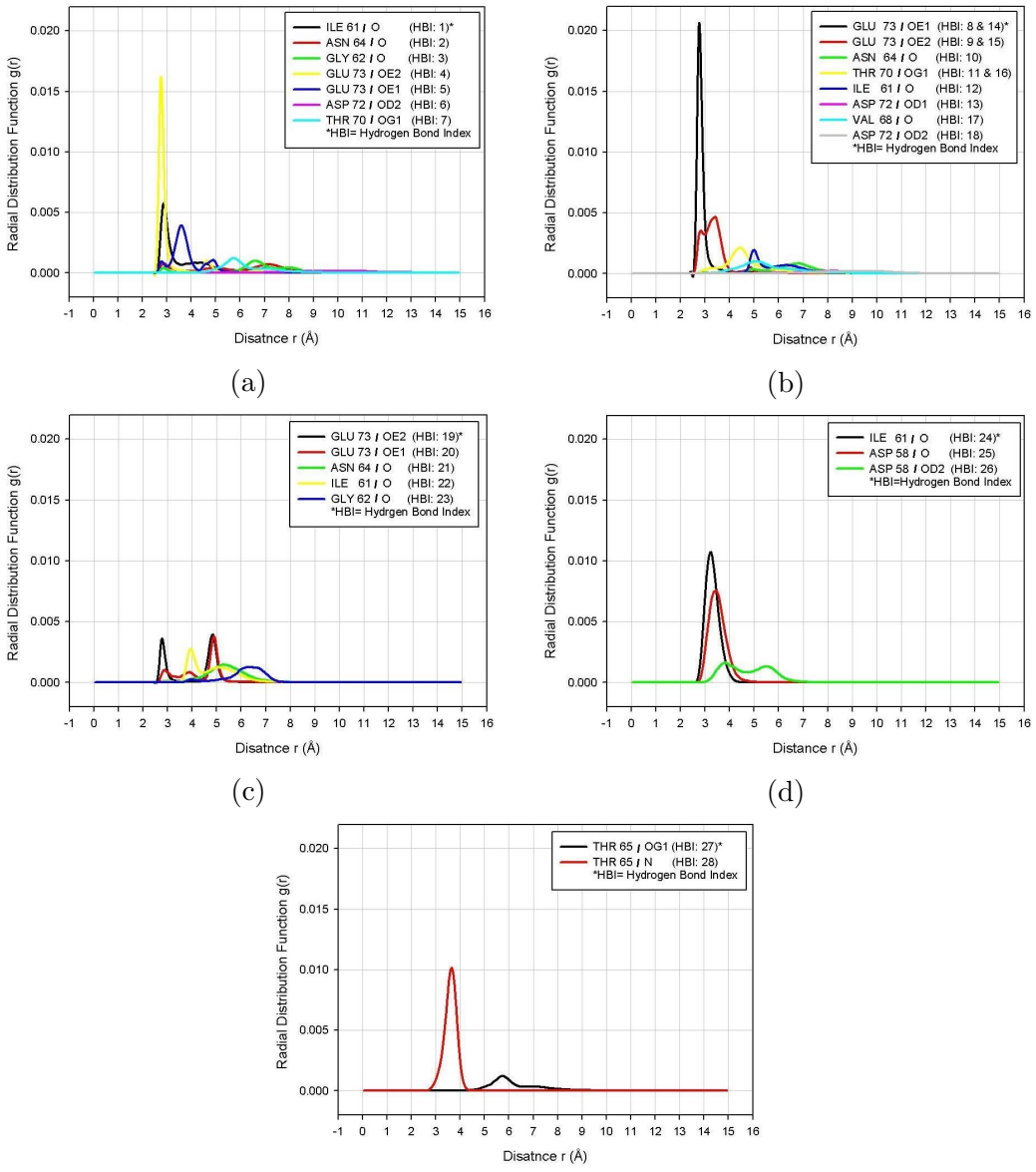
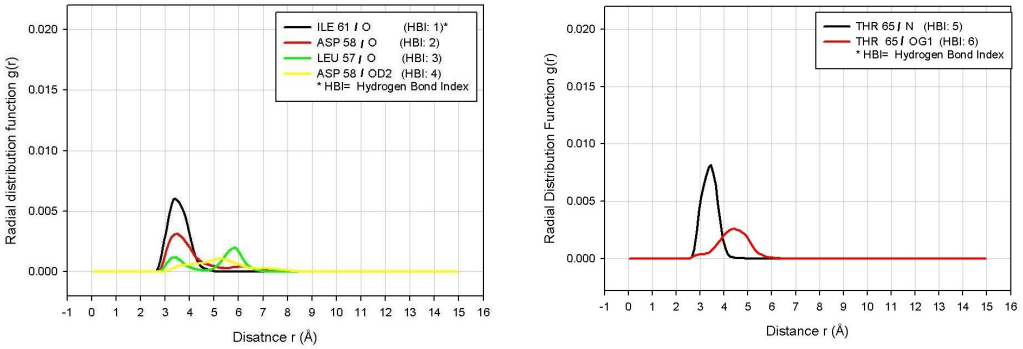


Figure 4.25: Radial distribution function of 262L model at 310K and ff99SB

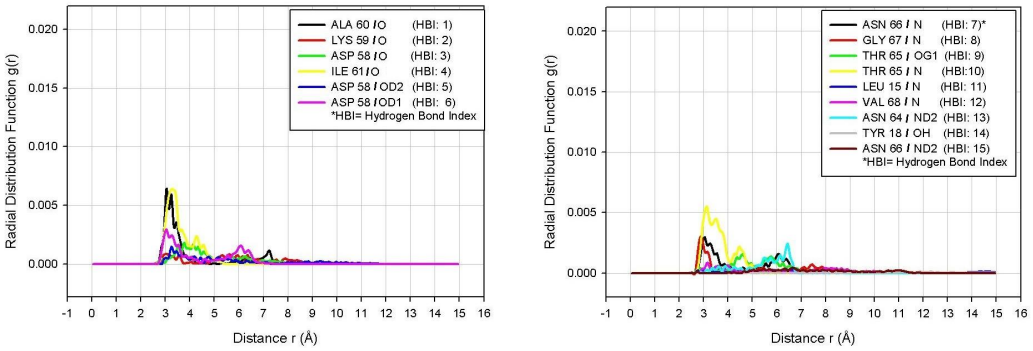
Radial distribution functions around donors and acceptors atoms in Arginine 63 including a)ARG 63/NH1, b)ARG 63/NH2, c)ARG 63/NHE, d)ARG 63/N and e)ARG 63/O with all corresponding acceptors and donors atoms from any other amino acid for all detected hydrogen bonds in table (4.2).



(a) (b)

Figure 4.26: Radial distribution function of 262L mutant model at 310K and ff99SB

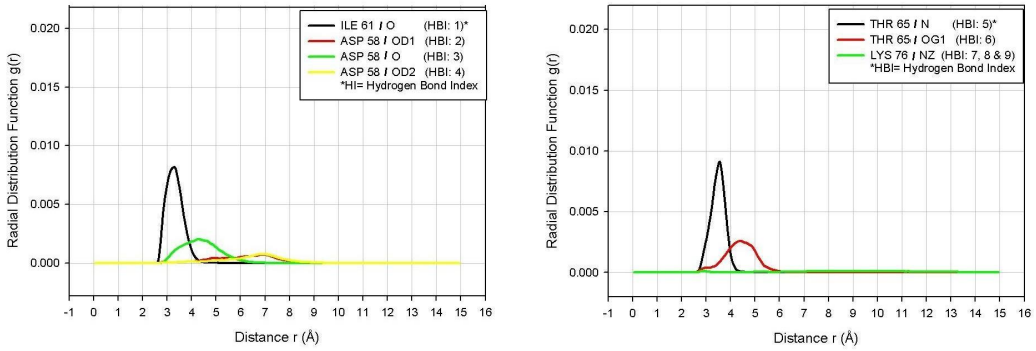
Radial distribution functions around donors and acceptors atoms in Alanine 63 including a)ALA 63/N and b)ALA 63/O with all corresponding acceptors and donors atoms from any other amino acid for all detected hydrogen bonds in table (4.3).



(a) (b)

Figure 4.27: Radial distribution function of 2F2Q model at 310K and ff99

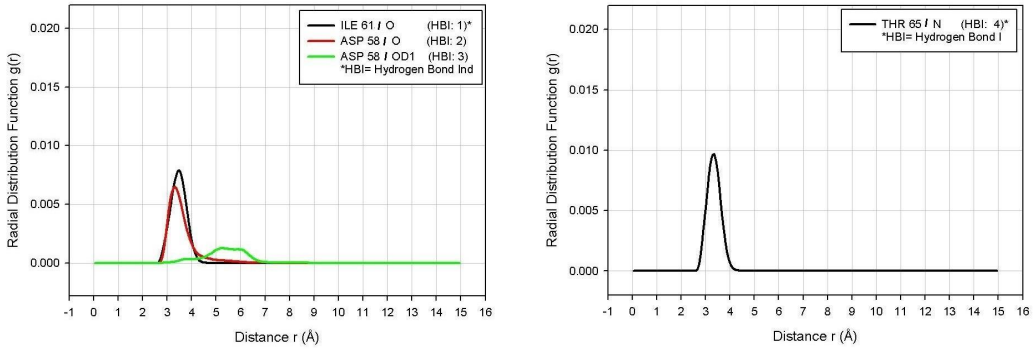
Radial distribution functions around donors and acceptors atoms in Alanine 63 including a)ALA 63/N and b)ALA 63/O with all corresponding acceptors and donors atoms from any other amino acid for all detected hydrogen bonds in table (4.4).



(a) (b)

Figure 4.28: Radial distribution function of 2F2Q model at 310K and ff99SB

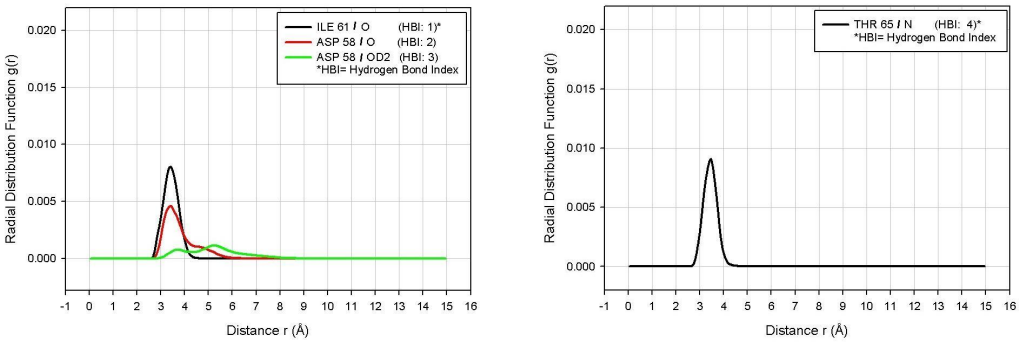
Radial distribution functions around donors and acceptors atoms in Alanine 63 including a)ALA 63/N and b)ALA 63/O with all corresponding acceptors and donors atoms from any other amino acid for all detected hydrogen bonds in table (4.5).



(a) (b)

Figure 4.29: Radial distribution function of 2F2Q model at 293K and ff99SB

Radial distribution functions around donors and acceptors atoms in Alanine 63 including a)ALA 63/N and b)ALA 63/O with all corresponding acceptors and donors atoms from any other amino acid for all detected hydrogen bonds in table (4.6).



(a) (b)

Figure 4.30: Radial distribution function of 2F2Q model at 277K and ff99SB

Radial distribution functions around donors and acceptors atoms in Alanine 63 including a)ALA 63/N and b)ALA 63/O with all corresponding acceptors and donors atoms from any other amino acid for all detected hydrogen bonds in table (4.7).

4.2 Discussion

Developing a Bio-molecular mechanical device has attracted the attention of several research groups due to its promising Nano and Bio-technological applications such as flow-control valves, switches and bio-sensors. An engineered model of T4 Lysozyme has been designed and reported for a new mechanical behavior, unrelated to the catalytic activity of the enzyme. A crystal structure of T4 Lysozyme is shown to prompt a large-scale translocation of the helix up to 20 $\mathring{A}$ upon insertion mutation of a helix sequence duplication of 39-50 amino acids followed by a point mutation of Arginine 63 to Alanine 63 (mutant L20/R63A). The duplicated-helix returned to its original conformation upon the addition of guanidinium ligand as a replacement for the guanidinium head group of the Arginine 63 side chain. This ligand seems to control the mechanical trans-location switching between two alter wild type on and off conformational states spanning over 20 $\mathring{A}$. In the on state, the N-terminal loop of the duplicated-helix is weakly structured, whereas the C-terminal loop has a well defined conformation that is stabilized with the Arginine 63 head group or the presence of the guanidinium ligand. In the off state, the C-terminal loop became destabilized and switches the protein by the trans-location of the duplicated helix towards the N-terminal due to the absence of the Arginine 63 head group in Ala 63 or the absence of the guanidinium ligand. The introduction of the guanidinium ligand can restore the protein from the off state to on state by re-stabilizing the C-terminal loop.

The purpose of this study is to investigate, in silico, the dynamical behavior and energetics of the engineered T4 Lysozyme as a step towards the development of a bio-molecular switch. Molecular dynamics simulations coupled with experimental investigation have been routinely used to study proteins structure function relationship and analyze underlying interaction between molecules. In addition, this study aims to build an initial mathematical model to describe motions observed by molecular dynamics simulations as a step towards understanding the physical basis and consequences of the large scale intrinsic hinge motion and/or the triggered motions.

Three approaches have been used in this study. The study started with one wild type crystal as a reference model for all simulations. 262L model has been used as the reference where Arginine is the amino acid at location 63 and the C-terminal loop is stabilized by means of non-bonding interaction with guanidinium head group of Arginine 63. New mutant 262L model with Arginine 63 to Alanine 63 mutation has been created, simulated in identi-

cal conditions and compared to the reference structure. This mutation was crested in order to probe the structural dynamics consequences of destabilizing the C-terminal loop. Another model, 2F2Q model, has Alanine as the amino acid of site 63 and C-terminal is stabilized by means of non-bonding interaction with the guanidinium ligand, mimicking the Arginine side chain. The guanidinium ligand was removed from the model to probe the structural dynamics consequences of destabilizing the C-terminal loop. Four identical copies of 2F2Q model have been created to be simulated at different conditions of force fields ,ff99 ff99SB, and temperatures of 4°C, 20°C and 37°C.

The first approach is to compare between the mutant R63A/262L model and the wild type 262l model. Mutating Arginine 63 to Alanine 63 destabilizes the C-terminal loop which was originally stabilized by means of non-bonding interaction with guanidinium head group of Arginine 63. Compared to the wild type, there are evident of loss in non-bonding interaction with amino acid 63 in the mutant R63A/262l model. Although the C-terminal loop was destabilized, the duplicated helix maintained its well defined structure and the triggered motion was not detected during 100ns of molecular dynamics simulation. Both models, the wild type and the mutant, have exhibited bending hinge motion during the course of simulation which suggests that this motion may be an intrinsic property of T4 Lysozyme independent of the mutation. However, the mutant 262L model has shown multiple hinges which may be because it shows better equilibrium behavior.

The second approach is to compare between two identical models 2F2Q, after removing the guanidinium ligand, at different conditions of force fields ,ff99 ff99SB, to the the reference model. Removing the guanidinium ligand from 2F2Q model destabilizes the C-terminal loop which was originally stabilized by means of non-bonding interaction with the guanidinium ligand. Compared to the wild type, there are evident of loss in non-bonding interaction with amino acid 63 for both models, however the duplicated helix maintained its well defined structure and the triggered motion was not detected during the course of simulation (100ns) with AMBER ff99SB force field. Interestingly, when using AMBER ff99 force field, initial stages of the triggered motion has been detected. One complete turn started to unfold from the N-terminal of the duplicated helix. Then the duplicated helix exhibited a slight shift ($\sim 7\mathring{A}$) towards the C-terminal. After that the entire protein became unstable before detecting the complete triggered $20\mathring{A}$ towards the C-terminal. AMBER ff99 force field [43,44] is a common force field for organic and bio-organic systems. It is an older version of AMBER ff99SB force field. It was reported to cause energy imbalance between the helical and the

extended regions of protein backbones and inaccurate treatment of glycine backbone parameters.[33] While AMBER ff99SB force field presents a careful reparametrization of the backbone torsion terms in ff99 and achieves much better balance of the four basic secondary structure elements.

The third approach is to compare between three models 2F2Q, removing the guanidinium ligand, at different conditions of temperatures of 4°C, 20°C and 37°C to the the reference model. Removing the guanidinium ligand stabilizes the C-terminal loop which was originally stabilized by means of non-bonding interaction with the guanidinium ligand. Compared to wild type, we found drop in non-bonding interaction with amino acid 63 for both models. Although the C-terminal loop was destabilized for all of the three models, the the duplicated helix maintained its well defined structure in and the triggered motion was not detected through the course of simulation (100ns). All models, at three different temperature, have exhibited bending hinge motion during the equilibration time which suggests that the motion the duplicated helix is an intrinsic property of T4 Lysozyme.

It appears as though for the triggered motion to be fully observed, the used force field will have to have somewhat flexible definition of the helical geometry without affecting the overall integrity of the protein. In other words, a hybrid force field where AMBER ff99 force field parameters are applied only to the duplicated helix and AMBER ff99SB force field on the rest of the protein could be useful in this situation accompanied with much longer simulation time and/or accelerated molecular dynamics.[51] It is very well known that conformation changes which involves folding and unfolding of protein might be in order of millisecond. [52]

On the other hand, all simulations constantly detect a large scale hinge motion (40°-90°) between the C- and N-terminal which did not involve folding and unfolding events. Dynamics of this helical hinge motion were detected and analysed. This behavior is in agreement with experimental studies reporting that in different crystal environments the lysozyme molecule displays a range of over 50° in the hinge bending angle between the N- and C-terminals. This conformational flexibility appears to be an intrinsic property of the protein, not due to the effect of different mutations.[14] More research is needed to study the possibility to control this helical hinge motion and/or trigger it and study its potential to be used in technological applications.

Chapter 5

The Mathematical Model

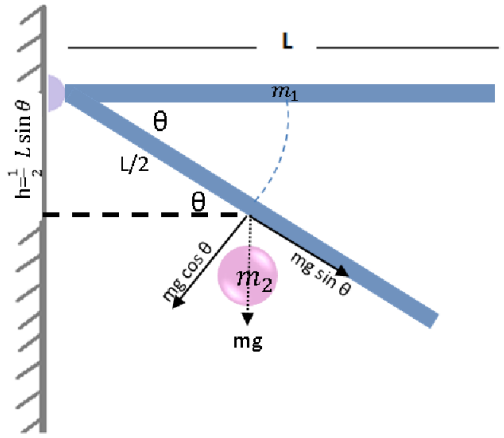


Figure 5.1: A pivoting rod

A preliminary mathematical model of the helical hinge motion as a pivoting rod. The engineered helix is considered as a thin rod of a mass m_1 and length L is released from rest in a horizontal position to rotate by angle about a friction less hinge as shown in figure (5.1). During rotation, the rod experiences a collision at its center of mass with a stationary object of mass m_2. This preliminary (solvent free) mathematical model is analogous to the detected hinge bending motion when the helix collides with a certain molecular object without taking into account the drag force of the surrounding solvent. The following physical picture is an attempt to model the rotating helix and calculate the impact force upon collision. First we should notice that every point in the rod, in particular, the center of mass is moving in as arc path as the rod rotates and experiences a centripetal acceleration.

For the center of mass, the path length equals

$$S = \frac{L}{2}\theta \tag{5.1}$$

and the centripetal acceleration is

$$a_c = m_1\frac{v_1^2}{r} = m_1\omega^2 r = m_1\frac{L}{2}\omega^2 \tag{5.2}$$

Where ω is the angular velocity and $r = \frac{L}{2}$ at the center of mass of the rod.

Before collision, in order to calculate the angular velocity, we must use the law of conservation of energy since the rod's angular acceleration is not uniform.

$$\Delta P.E. + \Delta K.E. = 0 \tag{5.3}$$

The change in potential energy basically depends on the vertical displacement traveled by the center of mass.

$$\Delta P.E. = -m_1 gh = m_1 g(\frac{-1}{2}L\sin\theta) \tag{5.4}$$

The change in kinetic energy is equal to the angular kinetic energy of the rod, since it starts from rest.

$$\Delta K.E. = \frac{1}{2}I\omega^2 = \frac{1}{2}(\frac{1}{3}m_1L^2)\omega^2 \tag{5.5}$$

By substitution from equation (5.4) and (5.5) in equation (5.3), law of conservation of energy becomes:

$$\frac{-1}{2}m_1gL\sin\theta + \frac{1}{2}(\frac{1}{3}m_1L^2)\omega^2 = 0 \tag{5.6}$$

$$g\sin\theta = \frac{1}{3}L\omega^2 = 0 \tag{5.7}$$

$$\omega = \sqrt{\frac{3g\sin\theta}{L}} \tag{5.8}$$

$$v_1 = \frac{L}{2}\sqrt{\frac{3g\sin\theta}{L}} \tag{5.9}$$

After collision, by applying the law of conservation of momentum

$$m_1v_1 = m_1v_1^{'} + m_2v_2^{'} \tag{5.10}$$

Where $v_1^{'}$and $v_2^{'}$ are the velocities after collision of the rod and the object respectively.

By applying law of conservation of energy assuming elastic collision

$$\frac{1}{2}m_1v_1^2 = \frac{1}{2}m_1v_1^{'2} + \frac{1}{2}m_2v_2^{'2} \tag{5.11}$$

Solving these two equations simultaneously, we got

$$v_1^{'} = v_1\frac{\mu\sqrt{-3\mu^2+4}}{2(\mu+1)} \tag{5.12}$$

$$v_1^{'} = \mu(v_1 - v_1^{'}) \tag{5.13}$$

Where $\mu = \frac{m_1}{m_2}$ is the masses ratio.

Knowing the velocity of the object after collision, one can calculate the average impact force exerted on the object due to collision with the rotating rod.

$$F_{average}\Delta t = m_2\Delta v_2 \tag{5.14}$$

$$F_{average} = m_2\frac{\Delta v_2}{\Delta t} = m_2\frac{(v_2^{'} - v_2)}{t} \tag{5.15}$$

Knowing the masses, the length of the helix, the angle of rotation and time of impact, we can calculate impact force excreted by the helix on any molecular object of interest upon collision.

Chapter 6

Conclusion

- Developing a Bio-molecular mechanical device has attracted the attention of several research groups due to its promising Nano and Biotechnological applications such as flow-control valves, switches and biosensors.
- An engineered model of T4 Lysozyme has been designed and reported for a new mechanical behavior, unrelated to the catalytic activity of the enzyme. A crystal structure of T4 Lysozyme is shown to prompt a large-scale trans-location of the helix up to 20 A upon insertion mutation α helix sequence duplication of 39-50 amino acids followed by a point mutation of Arginine 63 to Alanine 63 (mutant L20/R63A). The duplicated-helix returned to its original conformation upon the addition of guanidinium ligand as a replacement for the guanidinium head group of the Arginine 63 side chain.
- The purpose of this study is to investigate, in silico, the dynamical behavior and energetics of the engineered T4 Lysozyme as a step towards the development of a bio-molecular switch.
- Molecular dynamic simulations revealed a complex motion of an engineered surface helix within a protein molecule. At the Nanosecond time scale, Initial stages of triggered the motion that involve folding and unfolding of protein segments was partially observed about (7Å) after 80 ns of simulation only when the force field was relaxed for helical structures with the old AMBER ff99 force field.
- The duplicate helix of all protein models under investigation at different conditions of force fields and temperature exhibited hinge motion (33°-90°), during the course of simulation (100ns) which suggests that this motion is an intrinsic property of T4 Lysozyme.

- Advanced force fields combined with extended simulation time and/or accelerated molecular dynamics are needed to fully characterize the system.
- A preliminary (solvent free) mathematical model is constructed to theorize the mechanics of the helical hinge motion. It simulates the collision of the engineered helix with a molecular target to calculate the impact force. This model, upon further improvement, could be applicable for Nano and Bio-technological Applications.

Appendices

.1 Minimization Control File

1. imin=1 Flag to run minimization.
2. nmropt=0 No nmr-type analysis will be done.
3. ntmin=1 Flag for the method of minimization.
4. maxcyc=800 Steps of conjugate gradient.
5. ncyc=100 Steps of the steepest descent method.
6. dx0=0.0010 Initial step length.
7. drms=0.00010 Convergence criterion for the energy gradient.
8. ntx=1 Flag to read the initial coordinates and box size but not velocities.
9. irest=0 Do not restart the simulation; instead, run as a new simulation.
10. ntxo=1 Format of the final coordinates, velocities, and box size.
11. ntpr=100 Printing frequency of energy information.
12. ntrx=1 Format of the Cartesian coordinates for restraint.
13. ntwr=100 Printing frequency of the restart file.
14. iwrap=0 Do not wrap the coordinates written to the restart and trajectory files.
15. ntwx=100 Printing frequency of the coordinates in md coordinate file.
16. ntwv=0 Printing frequency of the velocities in md velocities file.
17. ntwe=0 Printing frequency of the velocities in md energies file.
18. ioutfm=0 Formatted ASCII trajectory for the format of coordinate and velocity trajectory files
19. ntwprt=0 Include all atoms of the system when writing trajectories.
20. idecomp=0 Do not perform energy decomposition.
21. ntf=2 Force evaluation and bond interactions involving H-atoms omitted.

22. ntb=1 Flag to assign constant volume periodic boundary conditions.
23. igb=0 Flag for not using the generalized Born or Poisson-Boltzmann implicit solvent model.
24. nsnb=1 Determines the frequency of nonbonded list updates.
25. cut=10.0 Flag to specify the nonbonded cutoff, in Angstroms.
26. ntc=2 Flag for SHAKE to perform bond length constraints on bonds involving hydrogen.
27. jfastw=0 Flag Fast water definition.
28. tol=0.00001 Relative geometrical tolerance for coordinate resetting in shake.
29. ntr=1 Flag for restraining specified atoms in Cartesian space using a harmonic potential.
30. restraint_wt=5.0 The weight (in kcal/mol -$\mathring{A}^2$) for the positional restraints.
31. restraintmask=$'$: $1-173'$ String that specifies the restrained atoms.

.2 Heating Control File

1. imin=0 Flag to run dynamics
2. nmropt=0 No nmr-type analysis will be done.
3. ntx=1 Flag to read the initial coordinates and box size but not velocities.
4. irest=0 Do not restart the simulation; instead, run as a new simulation.
5. ntxo=1 Format of the final coordinates, velocities, and box size.
6. ntpr=100 Printing frequency of energy information.
7. ntrx=1 Format of the Cartesian coordinates for restraint.
8. ntwr=100 Printing frequency of the restart file.
9. iwrap=0 Do not wrap the coordinates written to the restart and trajectory files.
10. ntwx=100 Printing frequency of the coordinates in md coordinate file.
11. ntwv=0 Printing frequency of the velocities in md velocities file.
12. ntwe=0 Printing frequency of the velocities in md energies file.
13. ioutfm=0 Formatted ASCII trajectory for the format of coordinate and velocity trajectory files
14. ntwprt=0 Include all atoms of the system when writing trajectories.
15. idecomp=0 Do not perform energy decomposition.
16. ntf=2 Force evaluation and bond interactions involving H-atoms omitted.
17. ntb=2 Flag to assign constant pressure periodic boundary conditions.
18. igb=0 Flag for not using the generalized Born or Poisson-Boltzmann implicit solvent model.
19. cut=10.0 Flag to specify the nonbonded cutoff, in Angstroms.
20. nstlim=25000 Number of MD-steps to be performed.

21. dt=0.002 The time step in units of (psec).
22. nscm=500 Flag for the removal of translational and rotational center-of-mass motion at regular intervals.
23. nrespa=1 This variable allows the user to evaluate slowly-varying terms in the force field less frequently.
24. ntc=2 Flag for SHAKE to perform bond length constraints on bonds involving hydrogen.
25. tol=0.00001 Relative geometrical tolerance for coordinate resetting in shake.
26. pres0=1.0 Reference pressure in units of (bars).
27. ntp=1 Flag for constant pressure dynamics with isotropic position scaling.
28. taup=2.0 Pressure relaxation time (in ps).
29. Tempi=100.0 Initial temperature in units of Kelvins (K).
30. Temp0=125.0 Reference temperature at which the system is to be kept at in units of Kelvins (K).
31. ntt=3 Use Langevin dynamics.
32. gamma_ln=2 The collision frequency γ, inps1.
33. ntr=1 Flag for restraining specified atoms in Cartesian space using a harmonic potential.
34. restraint_wt=5.0 The weight (in kcal/mol -$\mathring{A}^2$) for the positional restraints.
35. restraintmask=$'$: $1-173'$ String that specifies the restrained atoms.

.3 Molecular Dynamics Simulation Control File

1. imin=0 Flag to run dynamics.
2. nmropt=0 No nmr-type analysis will be done.
3. nstlim=25000 Number of MD-steps to be performed.
4. dt=0.002 The time step in units of (psec).
5. ntx=5 Coordinates and velocities will be read from a formatted (ASCII) coordinate file.
6. irest=1 Restart the simulation, reading coordinates and velocities from a previously saved restart file.
7. ntxo=1 Format of the final coordinates, velocities, and box size.
8. ntpr=100 Printing frequency of energy information.
9. ntrx=1 Format of the Cartesian coordinates for restraint.
10. ntwr=100 Printing frequency of the restart file.
11. iwrap=1 wrap the coordinates written to the restart and trajectory files.
12. ntwx=100 Printing frequency of the coordinates in md coordinate file.
13. ntwv=0 Printing frequency of the velocities in md velocities file.
14. ntwe=0 Printing frequency of the velocities in md energies file.
15. ioutfm=0 Formatted ASCII trajectory for the format of coordinate and velocity trajectory files.
16. ntwprt=0 Include all atoms of the system when writing trajectories.
17. idecomp=0 Do not perform energy decomposition.
18. ntb=2 Flag to assign constant pressure periodic boundary conditions.
19. igb=0 Flag for not using the generalized Born or Poisson-Boltzmann implicit solvent model.
20. cut=10.0 Flag to specify the nonbonded cutoff, in Angstroms.

21. ntr=0 Flag for not restraining any specified atom in Cartesian space using a harmonic potential.

22. nscm=500 Flag for the removal of translational and rotational center-of-mass motion at regular intervals.

23. nrespa=1 This variable allows the user to evaluate slowly-varying terms in the force field less frequently.

24. ntf=2 Force evaluation and bond interactions involving H-atoms omitted.

25. ntc=2 Flag for SHAKE to perform bond length constraints on bonds involving hydrogen.

26. tol=0.00001 Relative geometrical tolerance for coordinate resetting in shake.

27. pres0=1.0 Reference pressure in units of (bars).

28. ntp=1 Flag for constant pressure dynamics with isotropic position scaling.

29. taup=2.0 Pressure relaxation time (in ps).

30. Tempi=310.0 Initial temperature in units of Kelvins (K).

31. Temp0=310.0 Reference temperature at which the system is to be kept at in units of Kelvins (K).

32. ntt=3 Use Langevin dynamics.

33. gamma_ln=2 The collision frequency γ, inps1.

References

[1] Harvey Lodish, Arnold Berk & Paul Matsudaira. 1999. Molecular Cell Biology, Palgrave Macmillan; 4th edition, BkCdr.

[2] Berg JM, Tymoczko JL & Stryer L. 2002. Biochemistry. 5th edition, New York: W H Freeman.

[3] Anthony R. Poteete and Larry W. Hardy. 1994. Genetic Analysis of Bacteriophage T4 Lysozyme: Structure and Function. JOURNAL OF BACTERIOLOGY, 6783-6788.

[4] Streisinger, G., F.Mukai, W.J.Dreyer, B. Miller, and S. Horiuchi. 1961. Mutations affecting the lysozyme of phage T4. Cold Spring Harbor Symp.Quant.Biol, 26:25-30.

[5] Streisinger, G., Y.Okada, J.Emrich, J.Newton, A.Tsugita, E.Terzaghi, and M.Inouye. 1966. Frameshift mutations and the genetic code. Cold Spring Harbor Symp.Quant.Biol., 31:77-84

[6] Remington, S.J., W.F. Anderson, J. Owen, L F. Ten-Eyck, C.T. Grainger,and B. W. Matthews. 1978. Structure of the lysozyme from bacteriophage T4: an electron density map at 2.4 A resolution. J. Mol. Biol., 118:81-98

[7] Weaver, L H. & B.W. Matthews. 1987. Structure of bacteriophage T4 lysozyme refined at 1.7 A resolution. J. Mol. Biol., 193:189-199.

[8] Zhang, X.-J., Wozniak, J.A., Matthews, B.W. 1995. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. J. Mol. Biol, 250:527552.

[9] Bert L. de Groot, Steven Hayward, Daan van Aalten, Andrea Amadei & Herman J. C. Berendsen. 1998. Domain Motions in Bacteriophage T4 Lysozyme; a Comparison between Molecular Dynamics and Crystallographic Data. Proteins-Structure Function and Genetics., 31: 116-127.

[10] Strynadka NC & James MN. 1996. Lysozyme: a model enzyme in protein crystallography. EXS, 75:185-222.

[11] Chen, B.L., W. A. Baase, H. Nicholson, & J. A. Scheilman. 1992. Folding kinetics of T4 lysozyme and nine mutants at 12C. Biochemistry 31:1464-1476

[12] McIntosh, L .P., A.J. Wand, D.F. Lowry, A. G. Redfield, & F. W. Dahiquist. 1990. Assignment of the backbone'H and 15N resonances of bacteriophage T4 lysozyme. Biochemistry, 29, 6341-6362.

[13] Harpreet Kaur, Yellamraju U. Sasidhar. 2013. Molecular dynamics study of an insertion/duplication mutant of bacteriophage T4 Lysozyme reveals the nature of transition in full protein context. Physical Chemistry Chemical Physics, 15, 7819-7830.

[14] Zhang X-J, Wozniak JA & Matthews BW. 1995. Protein flexibility and adaptability seen in 25 crystal forms of T4 Lysozyme. Journal of Molecular Biology, 250,527-552.

[15] Jonathan W. Wray, Walter A. Baase, Joel D. Lindstrom, Larry H. Weaver, Anthony R. Poteete & Brian W. Matthews. 1999. Structural analysis of a non-contiguous second-site revertant in T4 lysozyme shows that increasing the rigidity of a protein can enhance its stability. Journal of Molecular Biology, 292, 5, 1111.

[16] Heinz DW., Baase WA. & Matthews BW. 1992. Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence. Proc Natl Acad Sci U S A., 89(9), 3751.

[17] Blaber M., Baase WA., Gassner N. & Matthews BW. 1995. Alanine scanning mutagenesis of the alpha-helix 115-123 of phage T4 lysozyme: effects on structure, stability and the binding of solvent. J. Mol. Biol., 246(2), 317.

[18] He MM1, Wood ZA, Baase WA, Xiao H & Matthews BW. 2004. Alanine-scanning mutagenesis of the beta-sheet region of phage T4 lysozyme suggests that tertiary context has a dominant effect on beta-sheet formation. Protein Sci., 13(10):2716-24.

[19] Sagermann, M., Baase, W.A., & Matthews, B.W. 1999. Structural characterization of an engineered tandem repeat contrasts the importance of

context and sequence in protein folding. Proc. Natl. Acad. Sci., 96: 6078-6083.

[20] Sagermann M., Gay L. & Matthews BW. 2003. Long-distance conformational changes in a protein engineered by modulated sequence duplication. Proc Natl Acad Sci U S A. 100(16):9191-5.

[21] Sagermann M., Baase WA., Mooers BHM, Gay L. & Matthews BW. 2004. Relocation or duplication of the Helix A sequence of T4 lysozyme causes only modest changes in structure but can increase or decrease the rate of folding. Biochemistry, 43:12961301

[22] Sagermann M., Baase WA., Matthews BW. 2006. Sequential reorganization of beta-sheet topology by insertion of a single strand. Protein Sci., 15 p.1085

[23] Weaver LH & Matthews BW. 1987. Structure of bacteriophage T4 lysozyme refined at 1.7 A resolution. J Mol Biol., 193(1):189-99.

[24] Kaur, H. & Sasidhar, Y. U., 2015, Environmental polarity indeces conformational transitions in a helical peptide sequence from bacteriophage T4 Lysozyme and its tandem duplicate: A molecular dynamics simulation study. Journal of Molecular Modeling, 21(4), 1-14.

[25] Yousef, M.S., Baase, W.A. & Matthews, B.W. 2004. Use of sequence duplication to engineer a ligand-triggered, long-distance molecular switch in T4 Lysozyme. Proc. Natl. Acad. Sci., 101, 11583-11586.

[26] Yousef, Mohammad, S.; Bischoff, Nicole; Dyer, Collin, M.; Baase, Walter, A. & Matthews, Brian, W. 2006. guanidinium derivatives bind preferentially and trigger long-distance conformational changes in an engineered T4 Lysozyme. Protein Science 15(4): 853-861.

[27] Ju Li. 2005. Basic Molecular Dynamics. Handbook of Materials Modeling, Springer. 565588.

[28] Grancharov D., Lilkova E., Ilieva N., Petkov P., Litov L., 2012, Open Problems in High-Performance Molecular-Dynamics Simulations, information technologies and control, 23-29

[29] David A. Case, Tom A. Darden, Tom E. Cheatham III, Carlos L. Simmerling, Kim Wang, Bob E. Duke, Ray Luo, Ross C. Walker, Wei Zhang, Kennie M. Merz, B. Roberts, Seth Hayik, Adrian Roitberg, Gustavo Seabra, Jason Swails, Andreas W. Gtz, Istvn Kolossvry, Kim F. Wong,

F. Paesani, Jiri Vanicek, Romain M. Wolf, Jian Liu, Xiongwu Wu, Scott R. Brozell, Thomas Steinbrecher, Holger Gohlke, Qin Cai, Xiang Ye, Jun Wang, M.J. Hsieh, Guanglei Cui, Dan R. Roe, Dave H. Mathews, M.G. Seetin, Romelia Salomon-Ferrer, Celeste Sagui, Volodymyr Babin, Tyler Luchko, Sergey Gusarov, Andriy Kovalenko, and Peter A. Kollman. 2012. AMBER 12, University of California, San Francisco.

[30] Dave A. Pearlman, David A. Case, Jim W. Caldwell, W.S. Ross, T.E. Cheatham III, Stephen E. Debolt, David M. Ferguson, George L. Seibel, and Peter A. Kollman. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comp. Phys. Commun., 91, 1-41.

[31] David A. Case, Tom E. Cheatham, Tom A. Darden, Holger Gohlke, Ray Luo, Kennie M. Merz, Alexey Onufriev, Carlos L. Simmerling, Bing Wang and R. Woods. 2005. The Amber biomolecular simulation programs. J. Computat. Chem., 26, 1668-1688.

[32] Romelia Salomon-Ferrer, David A. Case and Ross C. Walker. 2012. An overview of the Amber Biomolecular Simulation Package. WIREs Comput. Mol. Sci., 3(2), 198210.

[33] David A. Case, Dave A. Pearlman, Jim W. Caldwell, Tom E. Cheatham III, Ross C. Walker, Carlos L. Simmerling, Tom A. Darden, Kennie M. Merz, Robert V. Stanton, Ailan L. Cheng, James J. Vincent, Mike Crowley, David M. Ferguson, Randall J. Radmer, George L. Seibel, Chandra C. Singh, Paul K. Weiner, Peter A. Kollman, 2010. AMBER.

[34] Daniel R. Roe and Thomas E. Cheatham, III. 2013."PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data". J. Chem. Theory Comput., 9(7), 3084-3095.

[35] Humphrey, W., Dalke, A. & Schulten, K., 1996. "VMD - Visual Molecular Dynamics". J. Molec. Graphics, 14.1, 33-38.

[36] http://www.ks.uiuc.edu/Research/vmd/

[37] The PyMOL Molecular Graphics System, Version 1.7.4 Schrdinger, LLC.

[38] https://www.pymol.org/

[39] Steven Hayward & Herman J. C. Berendsen, 1998, "Systematic Analysis of Domain Motions in Proteins from Conformational Change; New Results on Citrate Synthase and T4 Lysozyme" Proteins, Structure, Function and Genetics, 30, 144.

[40] Steven Hayward, Akio Kitao & Herman J. C. Berendsen, 1997, "Model-Free Methods of Analyzing Domain Motions in Proteins from Simulation: A Comparison of Normal Mode Analysis and Molecular Dynamics Simulation of Lysozyme" Proteins, Structure, Function and Genetics, 27, 425.

[41] http://fizz.cmp.uea.ac.uk/dyndom/

[42] http://www.rcsb.org/pdb

[43] Wang, J.; Cieplak, P. & Kollman, P.A. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem., 21, 10491074

[44] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.& Roitberg, A.; Simmerling, C. 2006. Comparison of multiple Amber force fields and development of improved. Proteins, 65, 712725.

[45] Jorgensen W.L., Chandrasekhar J. & Madura J.D. 1983. Comparison of simplepotential functions for simulating liquid water. J. Chem. Phys., 79, 926935.

[46] Davis, M.& McCammon, J. 1989. Solving the finite-difference linearized Poisson-Boltzmann equation a comparison of relaxation and conjugate gradient methods. J. Comput. Chem., 10, 386391.

[47] Miyamoto, S.; Kollman & P. SETTLE. 1992. An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem., 13, 952962.

[48] Izaguirre, J.; Catarello, D.; Wozniak, J. & Skeel, R. 2001. Langevin stabilization of molecular dynamics. J. Chem. Phys., 114, 20902098.

[49] Pastor, R.; Brooks, B.& Szabo, A. 1988. An analysis of the accuracy of Langevin and molecular dynamics algorithms. Mol. Phys., 65, 14091419.

[50] Paulo Almeida. 2016. Proteins: Concepts in Biochemistry, Garland Science, Taylor and Fancies group, LLC

[51] Hamelberg, Donald; Mongan, John & McCammon, J. Andrew. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J. Chem. Phys., 2004, 120(24), 11919-11929.

[52] Lane TJ1, Shukla D, Beauchamp KA & Pande VS. 2013. To milliseconds and beyond: challenges in the simulation of protein folding. Curr Opin Struct Biol., 58-65.

List of Figures

List of Tables